218
27/8/10

THE AMERICAN JOURNAL OF PSYCHOLOGY

EDITED BY

KARL M. DALLENBACH
UNIVERSITY OF TEXAS

AND

M. E. BITTERMAN
BRYN MAWR COLLEGE

E. B. NEWMAN
HARVARD UNIVERSITY

LEO POSTMAN
UNIVERSITY OF CALIFORNIA, BERKELEY

WITH THE COÖPERATION OF

E. G. BORING, Harvard University; W. K. ESTES, Stanford University; J. P. GUILFORD, University of Southern California; HARRY HELSON, Kansas State University; E. R. HILGARD, Stanford University; FRANCIS W. IRWIN, University of Pennsylvania; G. L. KREEZER, Washington University; D. G. MARQUIS, Massachusetts Institute of Technology; W. C. H. PRENTICE, Wheaton College; T. A. RYAN, Cornell University.

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY

Vol. LXXV
MEZES HALL, THE UNIVERSITY OF TEXAS
AUSTIN, TEXAS

Copyright 1962 by Karl M. Dallenbach

Bur [illegible] ed [illegible] rch
27. 8. 70
J 213

TABLE OF CONTENTS

ARTICLES AND NOTES

ALLUISI, E. A., The Fifty-Fourth Annual Meeting of the Southern Society for Philosophy and Psychology 674
ANDREWS, J., and GOLDSTEIN, A. G., Perceptual Uprightness and Complexity of Random Shapes 667
ARNOULT, M. D., Ninth Annual Meeting of the Southwestern Psychological Association 675
AXELROD, S., and THOMPSON, L., On Visual Changes of Reversible Figures and Auditory Changes in Meaning 673
BAKAN, P., and THOMPSON, R., The Effect of Pre-Inspection Control Measures on the Size of Kinesthetic After-Effects 302
BAKAN, P., MYERS, L. B., and SCHOONARD, J., Kinesthetic After-Effects and Length of Inspection-Period 457
BARRATT, E. S., CLARK, M., and LIPTON, J., Critical Flicker Frequency in Relation to Cattell's Cultural Fair Intelligence Score 324
BARTZ, A. E., Electrical and Electronic Devices for the Teaching of Sampling Statistics 657
BAXTER, J. C., Mediated Generalization as a Function of Semantic Differential Performance 66
BECKER, S. W., and SIEGEL, S., Utility and Level of Aspiration 115
BEHREND, E. R., and BITTERMAN, M. E., Avoidance-Conditioning in the Goldfish: Exploratory Studies of the *CS-US* Interval 18
BEVAN, W., PRITCHARD, J. F., and REED, W. G., Single-Stimulus Judgments of Loudness as a Function of Presentation-Interval 612
BITTERMAN, M. E., and BEHREND, E. R., Avoidance-Conditioning in the Goldfish: Exploratory Studies of the *CS-US* Interval 18
BITTERMAN, M. E., and BULLOCK, D. H., Probability-Matching in the Pigeon 634
BITTERMAN, M. E., GONZALEZ, R. C., and MILSTEIN, S., Classical Conditioning in the Fish: Further Studies of Partial Reinforcement 421
BORING, E. G., The 1962 Meeting of the American Philosophical Society 327
BORING, E. G., The 1962 Meeting of the National Academy of Sciences 326
BRAUNSTEIN, M. L., An Automatic Apparatus for a Continuous Method of Limits 309
BROGDEN, W. J., Elmer Augustine Kurtz Culler: 1889-1961 155
BROOKS, V., and HOCHBERG, J., Pictorial Recognition as an Unlearned Ability: A Study of One Child's Performance 624
BROWN, D. R., MICHELS, K. C., and NAYLOR, J. C., Perception of Real Movement as a Function of Angle of Approach 144
BUCKHOUT, R., and NEWMAN, S. E., S-R and R-S Learning as Functions of Intralist Similarity 429
BUGELSKI, B. R., and HUFF, E. M., A Note on Increasing the Efficiency of Luchin's Mental Sets 665
BULLOCK, D. H., and BITTERMAN, M. E., Probability-Matching in the Pigeon 634
CAMPBELL, R. A., and SMALL, A. M., JR., Temporal Differential Sensitivity for Auditory Stimuli 401
CARDNO, J. A., The Experimental Upsurge, 1797-1874 499
CARLSON, V. R., Underestimation in Size-Constancy Judgments 462
CARLSON, V. R., and TASSONE, E. P., A Verbal Measure of the Perspective Attitude 644

CARRAN, A. B., Increasing Probability of Initial Recall with Additional Practice 466
CAZAVELAN, J., TREHUB, A., and TUCKER, I., Epidermal *b*-Waves and Changes in Basal Potentials of the Skin 140
CHAPLIN, M. R., and POLLACK, R. H., Some Methodological Problems in Microgenetic Investigation 495
CLARK, M., BARRATT, E. S., and LIPTON, J., Critical Flicker Frequency in Relation to Cattell's Cultural Fair Intelligence Score 324
COLLINS, C. C., JENSEN, A. R., and VREELAND, R. W., A Multiple S-R Apparatus for Human Learning 470
CORNSWEET, T. N., A Stabilized Image Requiring No Attachments to the Eye 653
CORNSWEET, T. N., The Staircase-Method in Psychophysics 485
CROVITZ, H. F., and ZENER, K., A Group-Test for Assessing Hand- and Eye-Dominance 271
CROWDER, W. F., A Simplified Constant-Current Shocking Circuit 306
DALLENBACH, K. M., An Acknowledgment 676
DATTA, L. -E. G., Learning in the Earthworm, *Lumbricus Terrestris* 531
DE WOLFE, A. S., Extinction of Conditioned Verbal Responses 265
DIGGORY, J. C., and OSTROFF, B., Estimated Probability of Success as a Function of Variability in Performance 94
DUFORT, R. H., The Psychological Laboratories at Wake Forest College .. 661
DUNCAN, C. P., Amount and Rate of Decay of Visual Figural After-Effect as Functions of Type of Inspection-Stimulus and Inspection-Time 242
DUNCAN, C. P., and LEWIS, D. J., Expectation and Resistance to Extinction under Partial Reinforcement and Risk-Taking 77
DYER, D. W., and HARCUM, E. R., Monocular and Binocular Reproduction of Binary Stimuli Appearing Right and Left of Fixation 56
EBENHOLTZ, S., and ROCK, I., Stroboscopic Movement Based on Change of Phenomenal Rather Than Retinal Location 193
EDWARDS, W., Measurement of Information Based on Over-Lapping Counts for Studying Sequential Dependencies 311
EPSTEIN, W., A Further Study of the Influence of Syntactical Structure on Learning 121
EYSENCK, H. J., WILLETT, R. A., and SLATER, P., Drive, Direction of Rotation, and Massing of Practice as Determinants of the Duration of the After-Effects from the Rotating Spiral 127
FARRIMOND, T., A Test of the Ability to Lip-Read—Use Visible Cues of Speech 477
FARROW, B. J., and SANTOS, J. F., A Note on Recency and Perception 671
FISHER, G. H., Phenomenal Causality in Conditions of Intrasensory and Intersensory Stimulation 321
FLOCK, H., and GIBSON, J. J., The Apparent Distance of Mountains 501
GALANTER, E., The Direct Measurement of Utility and Subjective Probability 208
GARVIN, E. A., and SANTOS, J. F., A Further Examination of the Schafer-Murphy Effect 259
GIBSON, E. J., PICK, A., HAMMOND, M., and OSSER, H., The Role of Grapheme-Phoneme Correspondence in the Perception of Words 554
GIBSON, J. J., and FLOCK, H., The Apparent Distance of Mountains 501
GOLDSTEIN, A. G., and ANDREWS, J., Perceptual Uprightness and Complexity of Random Shapes 667
GONZALEZ, R. C., BITTERMAN, M. E., and MILSTEIN, S., Classical Conditioning in the Fish: Further Studies of Partial Reinforcement 421
GREEN, E. J., and SQUIER, R. W., JR., The Rate of Learning a Discrimination in Relation to Some Predictive Measures 294
GRISWOLD, B. J., and LUCE, R. D., Choices among Uncertain Outcomes: A Test of a Decomposition and Two Assumptions of Transitivity 35
HAMMOND, M, GIBSON, E. J., PICK, A., and OSSER, H., The Role of Grapheme-Phoneme Correspondence in the Perception of Words 554
HARCUM, E. R., and DYER, D. W., Monocular and Binocular Reproduction of Binary Stimuli Appearing Right and Left of Fixation 56

HARTMAN, A., The Apparent Size of After-Images in Delusional and Non-Delusional Schizophrenics 587
HELM, C. E., and TUCKER, L. R., Individual Differences in the Structure of Color-Perception 437
HOCHBERG, J., and BROOKS, V., Pictorial Recognition as an Unlearned Ability: A Study of One Child's Performance 624
HOGAN, J. A., and ROZIN, P., An Improved Mechanical Fish-Lever 307
HUFF, E. M., and BUGELSKI, B. R., A Note on Increasing the Efficiency of Luchin's Mental Sets 665
HUNT, W. A., Carney Landis: 1897-1962 506
IVERSON, M. A., Thirty-Third Annual Meeting of the Eastern Psychological Association 504
JAHNKE, P. K., The Effects of a 'Compound Set' on Learning Efficiency 445
JENSEN, A. R., Temporal and Spatial Effects of Serial Position 390
JENSEN, A. R., COLLINS, C. C., and VREELAND, R. W., A Multiple S-R Apparatus for Human Learning 470
JOHNSON, R., and THOMSON, C., Incidental and Intentional Learning under Three Conditions of Motivation 284
JOHNSON, R. E., Qualitative Changes in the Memory-Trace of Incomplete Circles 629
JONES, F. N., and WOSKOW, M. H., On the Relationship between Estimates of Magnitude of Loudness and Pitch 669
JONES, F. N., SINGER, D., and TWELKER, P. A., A Waterless Thermal Stimulator 147
KAPPAUF, W. E., and YOUNG, P. T., Apparatus and Procedures for Studying Taste-Preferences in the White Rat 482
KENDLER, H. H., KENDLER, T. S., and LEARNARD, B., Mediated Responses to Size and Brightness as a Function of Age 571
KENNEDY, J. L., Fifty-Eighth Annual Meeting of the Society of Experimental Psychologists 328
KEPPEL, G., and UNDERWOOD, B. J., An Evaluation of Two Problems of Method in the Study of Retention 1
KEPPEL, G., REHULA, R., and UNDERWOOD, B. J., Item-Selection in Paired-Associate Learning 353
KINKADE, R. G., ROYCE, J. R., and STAYTON, W. R., Experimental Reduction of Autokinetic Movement 221
L'ABATE, L., Transfer of Learning with Differences in Associative Value and in Manifest Anxiety 251
LEARNARD, B., KENDLER, H. H., and KENDLER, T. S., Mediated Responses to Size and Brightness as a Function of Age 571
LEPLEY, W. M., An Apparatus for the Study of Binocular Rivalry 310
LEVIALDI, A., A Physical Model of Perception as an Integrating Process 325
LEWIS, D. J., and DUNCAN, C. P., Expectation and Resistance to Extinction under Partial Reinforcement and Risk-Taking 77
LIGHTFOOT, C., Contextual Influence upon the Perception of Color 503
LIPTON, J., BARRATT, E. S., and CLARK, M., Critical Flicker Frequency in Relation to Cattell's Cultural Fair Intelligence Score 324
LUCE, R. D., and GRISWOLD, B. J., Choices among Uncertain Outcomes: A Test of Decomposition and Two Assumptions of Transitivity 35
MCDONALD, R. P., Apparent Interposition in Binocular Depth-Perception 619
MCDONALD, R. P., An Artifact of the Brunswik Ratio 152
MCNEILL, D., and ROSENZWEIG, M. R., Inaccuracies in the Semantic Count of Lorge and Thorndike 316
MCNEILL, D., and ROSENZWEIG, M. R., Uses of the Semantic Count in Experimental Studies of Verbal Learning 492
MEDLIN, R. E., and SCOTT, T. R., Psychophysical Measurement of the Spiral After-Effect 319
MEYER, M. F., New Illusions of Pitch 323
MICHELS, K. C., BROWN, D. R., and NAYLOR, J. C., Perception of Real Movement as a Function of Angle of Approach 144

Miller, W. C., Thirty-Second Annual Meeting of the Rocky Mountain Psychological Association 505
Milstein, S., Bitterman, M. E., and Gonzalez, R. C., Classical Conditioning in the Fish: Further Studies of Partial Reinforcement 421
Moore, M. E., and Wallach, H., The Role of Slant in the Perception of Shape 289
Munson, R. F., Decision-Making in an Actual Gambling Situation 640
Myers, L. B., Bakan, P., and Schoonard, J., Kinesthetic After-Effects and Length of Inspection-Period 457
Naylor, J. C., Brown, D. R., and Michels, K. C., Perception of Real Movement as a Function of Angle of Approach 144
Newman, S. E., and Buckhout, R., S-R and R-S Learning as Functions of Intralist Similarity 429
Newman, S. E., and Saltz, E., Serial Positions as a Cue in Learning 102
Osser, H., Gibson, E. J., Hammond, M., and Pick, A., The Role of Grapheme-Phoneme Correspondence in the Perception of Words 554
Ostroff, B., and Diggory, J. C., Estimated Probability of Success as a Function of Variability in Performance 94
Oyama, T., The Effect of Hue and Brightness on the Size-Illusion of Concentric Circles 45
Peacock, L. J., and Williams, M., An Ultrasonic Device for Recording Activity 648
Pick, A., Gibson, E. J., Hammond, M., and Osser, H., The Role of Grapheme-Phoneme Correspondence in the Perception of Words 554
Poffenberger, A. T., Robert Sessions Woodworth: 1869-1962 677
Pollack, R. H., and Chaplin, M. R., Some Methodological Problems in Microgenetic Investigation 495
Poser, E. G., A Simple and Reliable Apparatus for the Measurement of Pain 304
Postman, L., Repetition and Paired-Associate Learning 372
Poulton, E. C., and Warren, R. M., Ratio- and Partition-Judgments 109
Pritchard, J. F., Bevan, W., and Reed, W. G., Single-Stimulus Judgments of Loudness as a Function of Presentation-Interval 612
Raab, D. H., Effect of Stimulus-Duration on Auditory Reaction-Time 298
Reed, W. G., Bevan, W., and Pritchard, J. F., Single-Stimulus Judgments of Loudness as a Function of Presentation-Interval 612
Rehula, R., Keppel, G., and Underwood, B. J., Item-Selection in Paired-Associate Learning 353
Roberts, W. A., A Further Test of the Effect of Isolation on Serial Learning 134
Rock, I., and Ebenholtz, S., Stroboscopic Movement Based on Change of Phenomenal Rather Than Retinal Location 193
Rosenzweig, M. R., and McNeill, D., Inaccuracies in the Semantic Count of Lorge and Thorndike 316
Rosenzweig, M. R., and McNeill, D., Uses of the Semantic Count in Experimental Studies of Verbal Behavior 492
Royce, J. R., Kinkade, R. G., and Stayton, W. R., Experimental Reduction of Autokinetic Movement 221
Rozin, P., and Hogan, J. A., An Improved Mechanical Fish-Lever 307
Saltz, E., and Newman, S. E., Serial Position as a Cue in Learning 102
Santos, J. F., and Farrow, B. J., A Note on Recency and Perception 671
Santos, J. F., and Garvin, E. A., A Further Examination of the Schafer-Murphy Effect 259
Sanua, V. D., Seventh Congress of the Interamerican Society of Psychology 328
Schoonard, J., Bakan, P., and Myers, L. B., Kinesthetic After-Effects and Length of Inspection-Period 457
Scott, T. R., and Medlin, R. E., Psychophysical Measurement of the Spiral After-Effect 319
Siegel, S., and Becker, S. W., Utility and Level of Aspiration 115
Singer, D., Jones, F. N., and Twelker, P. A., A Waterless Thermal Stimulator 147

Slater, P., Eysenck, H. J., and Willett, R. A., Drive, Direction of Rotation, and Massing of Practice as Determinants of the After-Effects from the Rotating Spiral 127
Small, A. M., Jr., and Campbell, R. A., Temporal Differential Sensitivity for Auditory Stimuli 401
Squier, R. W., Jr., and Green, E. J., The Rate of Learning a Discrimination in Relation to Some Predictive Measures 294
Stayton, W. R., Kinkade, R. G., and Royce, J. R., Experimental Reduction of Autokinetic Movement 221
Stevens, S. S., and Terrace, H. S., The Quantification of Tonal Volume .. 596
Tassone, E. P., and Carlson, V. R., A Verbal Measure of the Perspective Attitude 644
Terrace, H. S., and Stevens, S. S., The Quantification of Tonal Volume .. 596
Thompson, L., and Axelrod, S., On Visual Changes of Reversible Figures and Auditory Changes in Meaning 673
Thompson, R., and Bakan, P., The Effect of Pre-Inspection Control Measures on the Size of Kinesthetic After-Effects 302
Thomson, C., and Johnson, R., Incidental and Intentional Learning under Three Conditions of Motivation 284
Toch, H. H., The Effect of 'Meaning' on the Autokinetic Illusion 605
Trehub, A., Cazavelan, J., and Tucker, I., Epidermal *b*-Waves and Changes in Basal Potentials of the Skin 140
Tucker, I., Cazavelan, J., and Trehub, A., Epidermal *b*-Waves and Changes in Basal Potentials of the Skin 140
Tucker, L. R., and Helm, C. E., Individual Differences in the Structure of Color-Perception 437
Tunnell, E. H. B., A Bibliography of Articles and Books by Robert Sessions Woodworth: A Continuation, 1938-1959 690
Twelker, P. A., Jones, F. N., and Singer, D., A Waterless Thermal Stimulator 147
Umemoto, T., Paired-Associate Learning as a Function of Similarity: Semantic Similarity between Stimulus- and Response-Items 85
Underwood, B. J., and Keppel, G., An Evaluation of Two Problems of Method in the Study of Retention 1
Underwood, B. J., Keppel, G., and Rehula, R., Item-Selection in Paired-Associate Learning 353
Uttal, W. R., The Use of a Summary Card-Punch to Generate Stimuli and to Collect Data Simultaneously 150
Vreeland, R. W., Collins, C. C., and Jensen, A. R., A Multiple S-R Apparatus for Human Learning 470
Wallach, H., and Moore, M. E., The Role of Slant in the Perception of Shape 289
Wallach, L., The Complexity of Concept-Attainment 277
Warren, R. M., Are 'Autophonic' Judgments Based on Loudness? 452
Warren, R. M., and Poulton, E. C., Ratio- and Partition-Judgments 109
Waugh, N. C., Length of Series and the Learning-Curve 177
White, B. W., Stimulus-Conditions Affecting a Recently Discovered Stereoscopic Effect 411
Willet, R. A., Eysenck, H. J., and Slater, P., Drive, Direction of Rotation, and Massing of Practice as Determinants of the After-Effects from the Rotating Spiral 127
Williams, M., and Peacock, L. J., An Ultrasonic Device for Recording Activity 648
Woskow, M. H., and Jones, F. N., On the Relationship between Estimates of Magnitude of Loudness and Pitch 699
Wright, B., The Influence of Hue, Lightness, and Saturation on Apparent Warmth and Weight 232
Young, P. T., and Kappauf, W. E., Apparatus and Procedures for Studying Taste-Preferences in the White Rat 482

Zener, K., and Crovitz, H. F., A Group-Test for Assessing Hand- and Eye-Dominance 271

Book Reviews

(The reviewer's name appears in parentheses after the title of the work.)

Alexander H., Elements of Mathematical Statistics (J. Millman) 710
Andreas, B. G., Experimental Psychology (R. D. Walk) 170
Angona, L., *et al.*, L'Automation: Aspects Psychologiques et Sociaux (P. C. Smith) 706
Atkinson, R. C., and Suppes, P., Markov Learning Models for Multiperson Interactions (B. Fruchter) 707
Bailey, J. C., *et al.*, Organizational Behavior: Cases, Concepts and Research Findings (H. C. Triades) 520
Baker, L. M., General Experimental Psychology (R. D. Walk) 170
Barnes, L. B., *et al.*, Organizational Behavior: Cases, Concepts and Research Findings (H. C. Triades) 520
Barron, M. L., The Aging American: An Introduction to Social Gerontology and Geriatrics (N. W. Shock) 516
Berry, C. A., and Sells, S. B. (ed.), A Medical-Psychological Analysis (R. A. McFarland) 167
Birney, R. C., and Teevan, R. C., Instinct: An Enduring Problem in Psychology: Selected Readings (T. A. Ryan) 523
Birney, R. C., and Teevan, R. C., Reinforcement: An Enduring Problem in Psychology: Selected Readings (T. A. Ryan) 523
Birney, R. C., and Teevan, R. C., Color Vision: An Enduring Problem in Psychology: Selected Readings (T. A. Ryan) 523
Bonner, H., Psychology of Personality (M. H. Applezweig) 701
Books Received 525, 711
Brengelmann, J. C., and David, H. P. (ed.), Perspectives in Personality Research (S. Messick) 176
Brindley, G. S., Physiology of the Retina and the Visual Pathways (L. Hurvich and D. Jameson) 172
Brown, F., and Finn, M. H. P., Training for Clinical Psychology: Proceedings of the Springfield-Mt. Sinai Conference on Interim Training in Clinical Psychology (M. G. Clarke) 174
Brown, J. S., The Motivation of Behavior (R. C. Bolles) 699
Bühler, K., Das Gestaltprinzip im Leben des Menschen und der Tiere (E. Freeman) 338
Caldwell, W. E., *et al.* (ed.), Principles of Comparative Psychology (R. R. Zimmermann) 705
Chapanis, A., *et al.*, L'Automation: Aspects Psychologique et Sociaux (P. C. Smith) 706
Clark, J. V., *et al.*, Organizational Behavior: Cases, Concepts and Research Findings (H. C. Triades) 520
Clarke, A. D. B., and Clarke, A. M., Mental Deficiency: The Changing Outlook (W. Sloan) 340
Colby, K. M., An Introduction to Psychoanalytic Research (T. A. Ryan) 345
Coleman, J. C., Personality Dynamics and Effective Behavior (I. A. Berg) 708
Cooper, R., *et al.*, The Psychology of Character Development (H. H. Remmers) 334
Dalton, M., Men Who Manage: Fusions of Feeling and Theory in Administration (F. C. Mann) 165
David, H. P., and Brengelmann, J. C. (ed.), Perspectives in Personality Research (S. Messick) 176
Dawson, J. G., *et al.* (ed.), Psychotherapy and Schizophrenics: A Reappraisal (R. J. Silver) 710
Dellis, N. P., *et al.* (ed.), Psychotherapy and Schizophrenics: A Reappraisal (R. J. Silver) 710

DeSilva, E. R., and Dorcus, R. M., Science in Betting: The Player and the Horses (W. H. McGlothlin) 698
Doob, L., Becoming More Civilized: A Psychological Exploration (N. H. Pronko) 350
Dorcus, R. M., and DeSilva, E. R., Science in Betting: The Player and the Horses (W. H. McGlothlin) 698
Eysenck, H. J., Handbook of Abnormal Psychology: An Experimental Approach (N. H. Peters) 510
Feld, S., *et al.,* Americans View Their Mental Health: A Nationwide Interview Survey (E. S. Sulzer) 351
Field, J. (ed.), Handbook of Physiology: A Critical Comprehensive Presentation of Physiological Knowledge and Concepts. Section 1: Neurophysiology. Vols. I, II, III (H. S. Liddell) 693
Field, M. J., Search for Security: An Ethno-Psychiatric Study of Rural Ghana (M. H. Segall) 341
Finn, M. H. P., and Brown, F., Training for Clinical Psychology: Proceedings of the Springfield-Mt. Sinai Conference on Interim Training in Clinical Psychology (M. G. Clarke) 174
Fouraker, L. E., and Siegel, S., Bargaining and Group Decision Making Experiments in Bilateral Monopoly (J. L. Kennedy) 347
Frank, J. D., Persuasion and Healing: A Comparative Study of Psychotherapy (J. D. Matarazzo) 512
Galanter, *et al.,* Plans and the Structure of Behavior (T. A. Ryan) 161
Gubin, G., *et al.,* Americans View Their Mental Health: A Nationwide Interview Survey (E. S. Sulzer) 351
Hallman, R. J., Psychology of Literature: A Study of Alienation and Tragedy (A. H. Buchan) 169
Havighurst, R. J., *et al.,* The Psychology of Character Development (H. H. Remmers) 334
Hook, S., Psychoanalysis, Scientific Method, and Philosophy: A Symposium (E. C. Lambert) 518
Inhelder, B., *et al.,* The Child's Conception of Geometry (A. L. Baldwin) . 342
Iocono, G., *et al.,* L'Automation: Aspects Psychologiques et Sociaux (P. C. Smith) 706
Jackson, D. D., The Etiology of Schizophrenia (N. D. Sundberg) 514
Jacobson, E., *et al.,* L'Automation: Aspects Psychologiques et Sociaux (P. C. Smith) 706
Katona, G., The Powerful Consumer: Psychological Study of the American Economy (A. C. Freiberg) 522
Katz, R. L., *et al.,* Organizational Behavior: Cases, Concepts and Research Findings (H. C. Triades) 520
Lawrence, P. R., *et al.,* Organizational Behavior: Cases, Concepts and Research Findings (H. C. Triades) 520
Layton, W. L., The Strong Vocational Interest Blank: Research and Uses (L. H. Stewart) 350
Ledwith, N. H., A Rorschach Study of Child Development (F. S. Freeman) . 349
Lifton, W. M., Working with Groups (J. Harding) 351
Lilienthal, J., *et al.,* The Psychology of Character Development (H. H. Remmers) 374
Lucas, A., *et al.,* L'Automation: Aspects Psychologiques et Sociaux (P. C. (Smith) 706
Mackworth, N., *et al.,* L'Automation: Aspects Psychologiques et Sociaux (P. C. Smith) 706
McGuigan, F. J., Experimental Psychology: A Methodological Approach (D. L. Lordahl) 352
Meares, A., A System of Medical Hypnosis (G. H. Estabrooks). 344
Miller, G. A., *et al.,* Plans and the Structure of Behavior (T. A. Ryan) ... 161
Miller, G. E., (ed.), Teaching and Learning in Medical School (B. Kutner) 702
Miller, J. G., and Uhr, L., Drugs and Behavior (G. R. Wendt) 348

MISIAK, H. The Philosophical Roots of Scientific Psychology (R. C. MacLeod) 709
MOODIE, W., Hypnosis in Treatment (F. L. Marcuse) 352
MOWRER, O. H., Learning Theory and the Symbolic Process (D. E. Berlyne) 332
MOWRER, O. H., Learning Theory and Behavior (R. C. Gonzalez) 330
MURPHY, G., Challenge of Psychical Research: A Primer of Parapsyschology (J. L. Kennedy) 709
MURPHY, G., and SOLLEY, C. M., Development of the Perceptual World (J. F. Wohlwill) 163
ORTH, C. D., III, *et al.,* Organizational Behavior: Cases, Concepts and Research Findings (H. C. Triades) 520
PECK, R. F., *et al.,* The Psychology of Character Development (H. H. Remmers) 334
PEPPER, S. C., The Sources of Values (J. C. Diggory) 696
PIAGET, J., *et al.,* The Child's Conception of Geometry (A. L. Baldwin) 342
PRIBRAM, K. H., *et al.,* Plans and the Structure of Behavior (T. A. Ryan) 161
RETHLINGSHAFER, D. A., *et al.* (ed.), Principles of Comparative Psychology (R. R. Zimmermann) 705
SEBEOK, T. A., Style in Language (G. J. Suci) 351
SEILER, J. A., *et al.,* Organizational Behavior: Cases, Concepts and Research Findings (H. C. Triades) 520
SELLS, S. B., and BERRY, C. A. (ed.), A Medical-Psychological Analysis (R. A. McFarland) 167
SHEER, D. E. (ed.), Electrical Stimulation of the Brain: An Interdisciplinary Survey of Neurobehavioral Integrative Systems (B. H. Deatherage) 707
SIEGEL, S., and FOURAKER, L. E., Bargaining and Group Decision Making Experiments in Bilateral Monopoly (J. L. Kennedy) 347
SMITH, F. V., Explanation of Human Behaviour (R. W. Leeper) 336
SOLLEY, C. M., and MURPHY, G. Development of the Perceptual World (J. C. Wohlwill) 163
STONE, H. K., *et al.,* Psychotherapy and Schizophrenics: A Reappraisal (R. J. Silver) 710
SUPPES, P., and ATKINSON, R. C., Markov Learning Models for Multiperson Interactions (B. Fruchter) 707
SWANSON, G. E., The Birth of the Gods: The Origin of Primitive Beliefs (W. W. Lambert) 704
SZEMINSKA, A., *et al.,* The Child's Conception of Geometry (A. L. Baldwin) 342
TEEVAN, R. C., and BIRNEY, R. C., Instinct: An Enduring Problem in Psychology: Selected Readings (T. A. Ryan) 523
TEEVAN, R. C., and BIRNEY, R. C., Reinforcement: An Enduring Problem in Psychology: Selected Readings (T. A. Ryan) 523
TEEVAN, R. C., and BIRNEY, R. C., Color Vision: An Enduring Problem in Psychology: Selected Readings (T. A. Ryan) 523
TOMAN, W., An Introduction to Psychoanalytic Theory of Motivation (R. Littman) 346
TURNER, A. N., *et al.,* Organizational Behavior: Cases, Concepts and Research Findings (H. C. Triades) 520
UHR, L., and MILLER, J. G., Drugs and Behavior (G. R. Wendt) 348
VERNON, P. E., Intelligence and Attainment Tests (J. S. Ahmann) 343
VEROFF, J., *et al.,* Americans View Their Mental Health: A Nationwide Interview Survey (E. S. Sulzer) 351
WATERS, R. H., *et al.* (ed.), Principles of Comparative Psychology (R. R. Zimmermann) 705
WHYTE, L. L., The Unconscious before Freud (P. Madison) 345
WINN, R. B., Soviet Psychology (H. Pick) 173
WOLMAN, B. B., Contemporary Theories and Systems (R. W. Leeper) 336

THE AMERICAN
JOURNAL OF PSYCHOLOGY

THE AMERICAN
JOURNAL OF PSYCHOLOGY
Founded in 1887 by G. STANLEY HALL

Vol. LXXV MARCH, 1962 No. 1

AN EVALUATION OF TWO PROBLEMS OF METHOD IN THE STUDY OF RETENTION

By BENTON J. UNDERWOOD and GEOFFREY KEPPEL,
Northwestern University

This study deals with two problems of method which may arise in studying retention of verbal materials. One problem concerns the baseline from which to measure retention, the other concerns rehearsal of *S*s over long intervals of retention. The background of each problem will be considered in order.

The issues involved in determining the most appropriate baseline from which to measure forgetting can be best understood by considering a concrete problem. Let it be assumed that the purpose of an experiment is to measure retention as a function of meaningfulness, using two lists of items differing in meaningfulness. It is quite likely that the difference in meaningfulness will produce a difference in the rates at which the lists are learned. Therefore, the two lists must be carried to the same level of learning before the retention-interval is introduced; this is necessary that the measurements of retention will not be confounded by differences in degree of original learning. A common solution to this problem has been to carry the learning of each list to the same criterion, such as one perfect recitation. This solution, however, is imperfect since it can be shown logically and empirically that two lists learned at different rates to the same criterion do not exhibit the same degree of learning. Thus, if another trial were given on each list immediately after the criterion had been attained, differences in performance would be found. These differences would, therefore, confound measurements of retention taken later, *e.g.*, after 24 hr.

---

* Received for publication November 14, 1960. This work was supported by Contract Nonr-1228(15), Project NR 154-057, between Northwestern University and the Office of Naval Research.

The issue here should be clear. The difference in degree of learning attained when two lists are taken to the same criterion would not be great, and there is no reason to expect that the *rate* of forgetting would differ as a function of the small difference in degree of original learning. Even if the rate of forgetting does not differ, however, as a function of the small difference in degree of learning, the small difference is carried over intact to the measurements of retention. Differences in rate of forgetting as a function of the variable under investigation (meaningfulness, in the above illustration) are inferred from differences in the measurements. If these measurements include an effect due entirely to degree of learning in addition to an effect shown by the variable being investigated, it is quite clear that an inappropriate decision may be made concerning the influence of this variable. Differences in retention attributable to differences in degree of learning must be removed from the measurements if an accurate accounting of the effect of other variables on retention is to be obtained.

At least two solutions to the above problem have already been offered.

(1) One solution is given by using a projection of each *S*'s performance to an hypothetical next trial, the projection being based on group data.[1] Thus, these projections give an estimate of what the performance of each *S* would have been had there been another trial beyond the last one, and this projected score is used as the base from which to determine amount of forgetting over the interval of retention. This solution may be satisfactory, but it has one rather serious drawback; namely, the relatively large number of *S*s required to achieve stability in the projections. While there are no definitive data on the matter, the use of the method is not recommended unless a minimum of 30 *S*s is used as a base for the projections.

(2) A second solution to the problem is provided by the use of control groups, these groups being given immediate tests of retention. That is, these groups are taken to the criterion of learning established; then, following brief instruction for recall, they are given a test of retention. The elapsed time between the end of learning and recall is approximately 30 sec. These immediate values are then used as the base-values to determine the amount of forgetting shown by the experimental groups, *i.e.* groups having intervals of an hour, a day, or whatever interval has been chosen for the investigation. Thus, to continue the earlier illustration concerning meaningfulness, two control groups, C_1 and C_2, may show mean recalls of 8 and 9 items, respectively, and two experimental groups, E_1 and E_2, recalling after 24 hr., may recall 7 and 8 items, respectively. Since E_1-C_1 and E_2-C_2 are equal, the investigator would probably conclude that no difference in forgetting occurred. If the control groups had not been used, and if the difference between the recall of the two experimental groups was significant statistically, the investigator would probably have concluded that there was a true difference in rate of forgetting.

The use of control groups given immediate tests of recall would seem to be a near-perfect solution to the problem. It too, however, has potential deficiencies. Obviously, if a different control group is required for every experimental group, it is a costly procedure. There is also a possibility that the immediate-recall may introduce variables in performance that are not present when a test of retention is given after a long interval (*e.g.* 24 hr.). If *S* is warned before he starts learning that he will be stopped at some point and given a special test of retention, his learning-strategy may vary or he may be distracted by attempting to anticipate the point at which he will be stopped. If he is not warned ahead of time that he will be given a special test, he may be distracted during the test by puzzlement over the sudden change. It is not implied that any of these effects will be of a great magnitude even if they exist, but these minor effects may be important in the attempt to get a good estimate of forgetting when *S*s naïve to laboratory-studies are used. It is even conceivable that such effects may differ when the lists are presented under different conditions. For example, an immediate test of recall given to *S*s following learning by massed practice may be more disturbing than such a test given to *S*s following learning by distributed practice. The latter *S*s, in a manner of speaking, have had considerable practice in recalling after short intervals.

---

[1] Jack Richardson and B. J. Underwood, Comparing retention of verbal lists after different rates of acquisition, *J. gen. Psychol.*, 56, 1957, 187-192.

(3) A third solution to the problem of determining an accurate base for the measurement of forgetting is to use control groups which are given continuous learning-trials. Thus, rather than stopping the *S*s and requesting a special recall they would simply be given another learning-trial and performance on this trial would be used as the base from which amount of forgetting is calculated. This method would not diminish the number of groups required over that required by the control-recall method, except in the study of certain variables. For example, if one were to investigate retention as a function of number of learning-trials, a single control group could be used, this group being given one more learning-trial than the number given the experimental group having the highest degree of learning. The performance of this control group on any specified trial could be used as an estimate of how an experimental group would have performed had they not been stopped on the trial immediately preceding the specified trial. In general, this method would be more efficient than the control-recall method whenever the effect of a variable does not produce a difference in the rate of learning or a difference in the shape of the acquisition curve. Yet, if possible difficulties are introduced by the immediate-recall procedure, it might also be said that similar difficulties are present if the control performance is measured on a trial when the *S* was *not* specifically requested to recall. That is, it might be argued that since the experimental groups are given specific recall-instructions when they return to the laboratory after 24 hr., so too should the control groups.

It seems apparent that an empirical evaluation is needed of the comparability of these three methods for establishing a base from which forgetting is measured. Therefore, the first purpose of the present investigation is to compare base-measures estimated by these techniques: immediate recall; projection; and continuous learning. Since the accuracy of the estimates may vary as a function of degree of learning, comparisons will be made at three different degrees of learning.

The second problem with which the present study is concerned deals with *S*'s rehearsal during the intervals of retention, the duration of which is the period that *S* is out of the laboratory. Reports of *S* leave little doubt that such rehearsal occurs, and it occurs even if *E* attempts to mask the purpose of his return to the laboratory.[2] What effect such rehearsal has on retention is unknown. It might be suspected that any rehearsal would retard forgetting, but studies of rehearsal over short intervals are not at all clear on this matter.[3] Nevertheless, it would seem desirable, if possible, to prevent or minimize rehearsal. The second purpose of the present investigation is to evaluate the influence of instructions designed to prevent rehearsal.

METHOD

Lists. The two lists of eight paired associates used as the learning-material are shown in Table I. The stimuli were single letters, the responses two letters. These lists are heterogeneous lists in the sense that the associative connection between the stimulus-letter and the first letter of the response (S-1) is either high or low, and within these two categories four degrees of associative connection exist between the first two letters (stimulus-letter and first response letter) and the third letter (S1-2). The values in Table I represent an index of the strength of associative connection as taken from the Underwood-Schulz tables.[4] The higher the value the greater is the associative connection.

TABLE I

LISTS USED, WITH INDEX OF ASSOCIATIVE CONNECTION BETWEEN THE STIMULUS-LETTER AND FIRST LETTER OF THE RESPONSE (S-1), AND BETWEEN THE FIRST TWO LETTERS AND SECOND LETTER OF THE RESPONSES (S1-2)

List 1	S-1	S1-2	List 2	S-1	S1-2
N-OP	170	43	A-ST	39	76
U-TE	41	25	X-IL	50	28
I-MS	31	15	W-EM	40	17
J-KX	26	1	D-OJ	130	0
G-RA	5	61	B-QR	0	48
L-CD	3	24	F-VN	2	29
Z-FB	3	15	P-HY	1	11
Q-HV	1	3	K-ZC	2	1

The two lists have essentially the same structure; two lists were used in the interests of generality and no evaluation of differences between the lists will be made. Each list was used equally often with each condition. It will be noted that no letter is repeated within either of the lists. The lists were presented at 2:2-sec. rate with a 4-sec. intertrial interval. Four different orders of presentation were employed and each order was used equally often as a start-order.

Immediate-recall conditions. Immediate recall was inserted at three different levels of learning, and since a given group had only one such recall-test, the point of immediate recall can be used to identify the groups. Group 4 was given immediate-recall instructions after Trial 4, Group 12 after Trial 12, and Group 24, after Trial 24. Therefore, the effect of the immediate-recall instructions may be measured on trials 5, 13, and 25, respectively. Groups 4 and 12 were given a total of 13 acquisition-trials; Group 24 had 25 acquisition-trials. Finally, there was a fourth group given 25 continuous learning-trials; this group will be referred to as Group N (no immediate recall). Each of these four groups consisted of 32 *Ss*, all naïve to verbal-learning experiments.

---

[2] B. J. Underwood and Leo Postman, Extraexperimental sources of interference in forgetting, *Psychol. Rev.*, 67, 1960, 73-95.

[3] A. B. Shaklee and B. E. Jones, Problems in method and theory in controlling rest activity, *J. gen. Psychol.*, 60, 1959, 11-16.

[4] B. J. Underwood and R. W. Schulz, *Meaningfulness and Verbal Learning*, 1960, Appendix F.

For immediate recall, *E* stopped the memory-drum, walked around to *S*'s side of the screen, and read the following:

> I have stopped the drum to give you a moment's rest. However, we are going to go on with the learning immediately. When I start the drum I want you to give as many correct responses as possible on the very first trial and to continue to anticipate as many as possible on each trial until I stop you.

Following these instructions, *E* immediately returned to his chair, said 'Ready,' and started the drum. On the average, 30 sec. elapsed between the time *E* stopped the drum and started it again.

Rehearsal-conditions. As a consequence of the above procedures, 64 *Ss* were given 13 acquisition-trials and 64 were given 25 acquisition-trials. After the requisite number of trials was completed, half the *Ss* is each group were given instructions designed to prevent rehearsal during the 24-hr. interval of retention, and half were simply reminded to return the following day at the same time for additional work. These latter groups will be called the NI Groups (not instructed).

The instructed groups (I Groups), after completing original learning, were read the following instructions:

> As you know, you are to return tomorrow at the same time for the next part of the experiment. We want you to watch out for something between now and then. Between now and tomorrow we don't want you to think about the list of letters that you have just learned. The reason for this is that your data tomorrow will be much less useful to us if you do allow yourself to think about the list between now and then. To *not* think about the list may not be as easy as it first seems; many people have told us that after learning a list like this the items keep popping into their heads, and for no particular reason. While we don't know of any way to keep them from popping in, we do know a way to push them out again quickly. When any of the items pops up, all you have to do is to try to think of something else—anything else. Will you please try to do this; that is, if you do have a tendency to think about the items, the moment this happens start thinking about something else. In this way your work for tomorrow will be of maximal value to us.

The division of the original groups (Groups 4, 12, 24, and N) into NI and I Groups was, of course symmetrical. That is, half of the 32 *Ss* in each of the four groups became members of I Groups and half became members of NI Groups.

When *S* returned to the laboratory 24 hr. after original learning, recall-instructions were read and five relearning-trials were given. The recall instructions stressed the importance of getting as many correct responses on each trial as possible.

RESULTS

Acquisition. Learning-curves for the four groups are shown in Fig. 1. In assigning the *Ss*, a random order of the four conditions was made up, and they were assigned to this order in sequence as they appeared at the laboratory. In spite of this random assignment, the differences in acquisition appear somewhat larger than would be anticipated. However, the *F*s for the total correct responses for the first 4 trials and for the first 13 trials for the four groups do not attain the 5% level of significance. Nevertheless, the differences in performance are large enough to recommend that certain selections be made in obtaining values for evaluating differences in the three methods for estimating base retention measurements. First, however, some evaluation of item-differences in learning will be made.

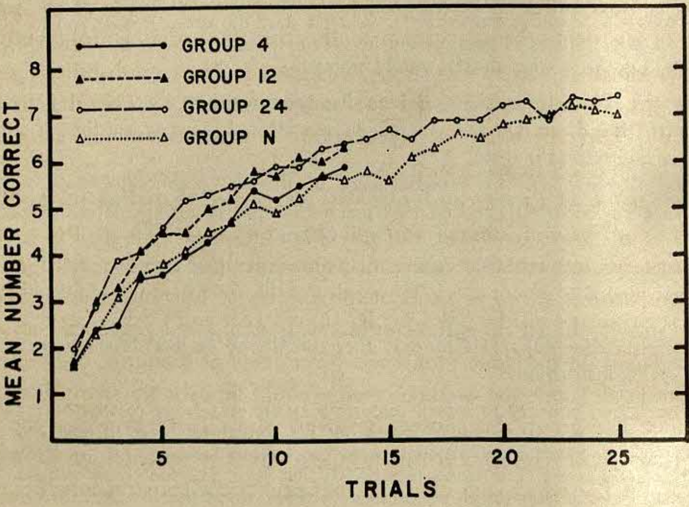


FIG. 1. ACQUISITION-CURVES
The number identifying each group indicates the trial after which instructions for immediate recall were given (*e.g.* Group 4 was given such instructions after Trial 4). Group N did not receive an immediate test of recall.

Associative connection and learning. A simple index of the over-all strength of preëxisting associative connections between letters for a given item was obtained by summing the two values for each item as given in Table I. Thus, N-OP was given a value of 213, U-TE 66, and so on. How do these summed values relate to learning-scores? Each list was presented to 32 *S*s for 13 trials and for a like number (but different) *S*s for 25 trials. For each item in the lists the total number of times that each response was anticipated correctly was determined for the 13-trial *S*s and also for the 25-trial *S*s. For the 13-trial *S*s, the range of total correct responses for the 16 items for 32 *S*s extended from 369 (N-OP) to 94 (K-ZC). These same two items were also the easiest and most difficult, respectively, for the 25-trial *S*s. The product-moment correlation between the summed associative index and learning for the 13-trial *S*s was 0.82; for the 25-trial *S*s the value was 0.77. The *r* between the learning scores by items for the 13-trial *S*s and the 25-trial *S*s was 0.97. The only item which showed a serious discrepancy between the associative index and learning was P-HY; this item was learned much more rapidly than its associative index would predict.

For another index of learning, a determination was made of the probability of an item's being correct on the trial immediately following the trial on which it was first correctly anticipated. These probabilities (combining the 13-trial and 25-trial *S*s) ranged from 0.83 (N-OP) to 0.35 (G-HV). The *r* between these probability-values and summed associative index for the 16 items was 0.69.

Comparison of base-measures. It will be recalled that three methods for determining base-measures were to be compared. These three base-measures are immediate-recall, continuous-learning, and projected scores. The comparisons are to be made at three levels of learning; namely, Trials 5, 13, and 25. In all cases the continuous-learning and projected measures will be based on the same group at a given level of learning. For comparisons on Trial 5 any one of three groups could be used to obtain these two measures. The group was chosen, however, whose performance on the first four trials most closely corresponded to the performance of Group 4 (the group for which the effect of immediate-recall instructions between Trials 4 and 5 will be determined). An inspection of Fig. 1 will show that this group is Group N. For comparisons on Trial 13, Group 24 is used for the continuous and projected measures because the performance of this group through the first 12 trials most closely corresponded to Group 12 (given immediate-recall instructions following Trial 12). Obviously, for the 25-trial comparisons, Group N must be compared with Group 24.

For all comparisons of measures, change-scores have been used. These change-scores were determined for each *S*. For example, consider the comparisons to be made on Trial 5. For the immediate-recall measure (Group 4) the difference in the performance of each *S* on Trial 4 and Trial 5 was determined. For most *S*s at this level of learning the change-scores are positive, *i.e. S* does better on Trial 5 than on Trial 4. For the continuous-learning measure (Group N), exactly the same procedure was used. The determination of the projected measure is, however, somewhat more complicated and will be explained in detail, using Trial 5 as an illustration.

As noted above, to derive the projected measure for Trial 5, Group N was used. The first step involved determining the percentage of items correct on Trial 4 as a function of the number of times they were given correctly (0, 1, 2, or 3 times) during the first 3 anticipation-trials. That is, Trial 4 is the observational trial from which the growth of associative strength as a function of number of prior correct anticipations is determined. Since there were 32 *S*s and 8 items for each, a total of 256 items was available for this determination. These 256 items were distributed as follows: 138 were not correctly anticipated during the first three trials, and 48, 36, and 34 items were correctly anticipated 1, 2, and 3 times, respectively. On Trial 4 the percentages correct were 17, 71, 78, and 97 following 0, 1, 2, and 3 correct anticipations, respectively. Obviously, these values indicate that the greater the number of times an item has been anticipated correctly on the first 3 trials the greater is the probability that it will be correct on Trial 4.

The next step was to estimate a score for each *S* for an hypothetical Trial 5, based on information available from the first 4 trials. The original tally does not give a probability-estimate for an item's being correct on Trial 5 following 4 correct anticipations on the first 4 trials. A close approximation can be made, however, for such an item by merely extrapolating the probability-curve based on 0, 1, 2, and 3 correct anticipations. For the present data this has been estimated as 0.99; that is, if, on 4 anticipation-trials, an item has been given correctly 4 times, it is estimated that its probability of being correct on Trial 5 is 0.99. It is assumed, then, that items given correctly 0, 1, 2, 3, and 4 times on the first 4 trials will have probabilities of being correct of 0.17, 0.71, 0.78, 0.97, and 0.99, respectively, on Trial 5.

Based on these group 'norms' for all *S*s for all items, an expected or projected score for Trial 5 can be obtained for each *S*. For example, *S* No. 115 had 4 items which were never correctly anticipated during the first 4 trials; 1 item was correctly anticipated 3 times, and 3 items were correctly anticipated 4 times. The projected score for Trial 5 is obtained by summing the group-probabilities for individual items; namely, 0.17, 0.17, 0.17, 0.17, 0.97, 0.99, 0.99, and 0.99. The sum is 4.62; thus, the projected score is 4.62 items. This particular *S* had 4 items correct on Trial 4; since the predicted value for Trial 5 is 4.62 the change-score is 0.62.

Such change-scores were obtained for each *S*; the mean of this distribution of change-scores may be compared with the means of the distributions of change-scores for immediate recall and for continuous learning. All three sets of change-scores have been determined for each of the three levels of learning, and are shown graphically in Fig. 2.

All statistical evaluations of differences apparent in Fig. 2 have been made by *t*-ratios. Since the projected and continuous measures are based on the same *S*s, the direct-difference *t* has been used for comparing these measures. When the immediate-recall scores are compared with either of the other two measures, the random-groups *t* is used. The results of the *t*-tests are definitive. None of the differences between any two measures for the change-scores between Trials 12 and 13 yield a *t* greater than 1; the same holds true for differences in mean change-scores between Trials

24 and 25. For example, although Fig. 2 may suggest that the difference between continuous learning and immediate recall for Trial 13 is rather large, the *t* for the difference between the means is only 0.94. The mean change estimated by the projected measure is, however, significantly higher than either of the other two measures for the change between Trials 4 and 5. At this point in learning the projected vs. continuous measure yields a *t* of 2.10 which, with 31 *df* is significant at the 5% level. The projected vs. recall measure yields a *t* of 2.50. With 62 *df* the difference is just short of the 1% level of significance. Thus, it is clear that at the low degree of learning the projected measure gives a higher estimate of performance than do the other two measures. At the two other levels of learning, however, the three measures give essentially the same estimate.

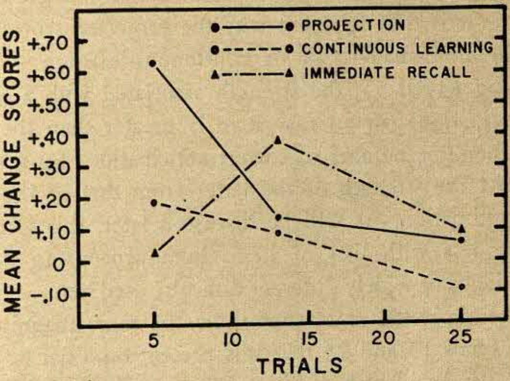


FIG. 2. COMPARISON OF CHANGE-SCORES BETWEEN TRIALS 4 AND 5, 12 AND 13, AND 24 AND 25
(As estimated by projection, immediate recall, and continuous learning.)

The failure of the projected score to predict the performance on Trial 5 adequately is probably to be attributed to the wide differences in difficulty of the items in the list. The projected score is based on the assumption that the obtained probability of being correct on Trial 4 (as a function of a given number of correct anticipations on Trials 1 through 3) will predict accurately the probabilities of being correct on Trial 5 following the same given number of correct anticipations on Trials 1 through 4. Thus, in obtaining the predicted probability, two correct anticipations on Trials 1 through 3 were assumed to produce the same associative strength as two correct anticipations on Trials 1 through 4. If, however, the items within a list differ widely in difficulty, this assumption may not be valid. An item given correctly twice in three trials would be a less difficult item than one given correctly twice in four trials, and to use the associative strength of the former to predict the associative strength of the latter should produce an overestimation of associative strength. An empirical check has been made on this likelihood by determining the actual probabilities of being correct on Trial 5 as a function of the number of correct anticipations on Trials 1 through 4; these probabilities were compared with those determined for Trials 1 through 3. These comparisons showed that the obtained probabilities were higher in every category on Trial 4 than on Trial 5 except for zero correct anticipations, where the probabilities were equal. It is thus apparent why the projected scores were higher than either the immediate-recall or continuous-learning scores. At this level of learning (Trial 5), the strength associated with a given number of correct anticipations for an easy item is greater than the strength associated with the same number of correct anticipations for a more difficult item. To predict the strength of the latter from that of the former produces an overestimation. As will be discussed later, this error of estimation should be less with lists of items homogeneous in difficulty than for the lists of items of widely different difficulty used here.

Why did the projected scores not show the same magnitude of overestimation on Trials 13 and 25 that was present on Trial 5? Three interrelated factors may be identified. First, differences in probabilities of items differing widely in *initial* difficulty are relatively small after several correct anticipations on each. Secondly, at the levels of learning represented on Trials 13 and 25, most items have been correctly anticipated at least once so that probabilities are based on items representing the full range of difficulty. Finally, the greater the number of trials, the more equal the difficulty of the predictor and predicted items. That is to say, the difference in difficulty of two items correctly anticipated two and three times, respectively, on the first three trials is greater than the difference in difficulty of two items anticipated two and three times, respectively, on the first 11 trials. All of these three factors would tend to make the projected scores more closely approximate both the immediate-recall and continuous-learning scores on Trials 13 and 25 than was true on Trial 5.

Anti-rehearsal instructions and retention. It will be remembered that 64 *S*s were given 13 learning-trials and 64 were given 25 trials. Half of the *S*s in each group were given instructions designed to prevent rehearsal (I Groups) and half were not (NI Groups). Did retention vary as a consequence of the instructional differences? It will be seen that the data do not provide a simple answer to this question.

The last five original learning (OL) trials and the five relearning-trials for each level of learning are plotted in Fig. 3. Attention will be directed first to retention of the groups having 13 OL trials. It will be noted that the two groups differ consistently both in OL and in relearning. Since the two groups were treated exactly the same during OL, the differences must be attributed to sampling-variations. The implication of this fact is simply that direct comparisons cannot be made of the recall-scores since these scores undoubtedly include a difference present at the end of learning. Three independent analyses have been made to adjust for the differences in original learning. The first method involves the use of projected scores (as explained earlier). For the I Group the mean projected score is 6.23, a value closely approximating that obtained if the original learning-curve is extrapolated to a 'next trial.' For the NI Group, the projected value is 6.54. Loss-scores over 24 hr. were calculated for each *S* and a *t* between the means of the two groups determined. This value is 2.14; with 62 *df*, a value of 2.00 is signficant at the 5% confidence-level.

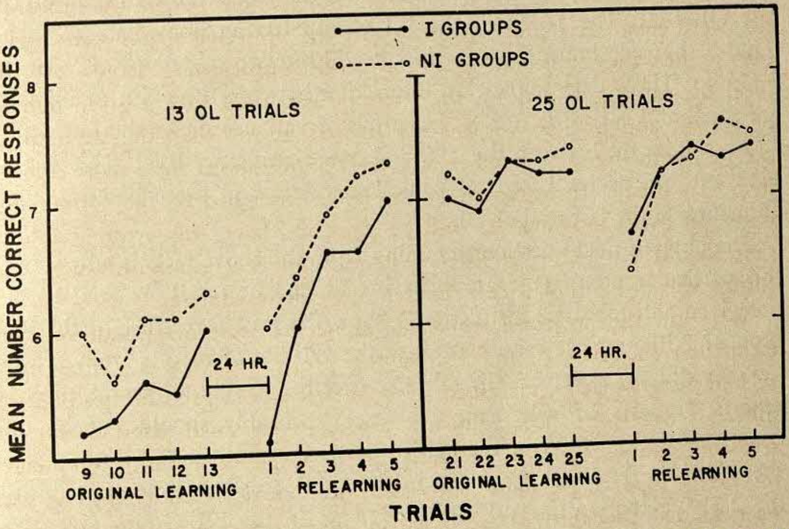


FIG. 3. RETENTION AS A FUNCTION OF ANTI-REHEARSAL INSTRUCTIONS (Group I) VERSUS NO INSTRUCTIONS (NI Groups) AT TWO LEVELS OF LEARNING

For the second analysis the loss for each *S* over the 24-hr. interval was determined by subtracting his recall-score from the score made on the last learning trial. The *t* for the mean difference (0.83) of the loss-scores was 1.85, a value which is a little short of the 5% level of significance.

Application of analysis of covariance, which might be judged to be the most appropriate analysis, gives an *F* of 5.31. An *F* of 4.00 is needed for the 5% significance-level.

These analyses lead to the conclusion, but by no means a firm conclusion, that the group not given instructions designed to prevent rehearsal did not forget as much as did the instructed group. It will be remembered that half of the *S*s in each group were given a test of immediate recall on Trial 5 and half were given such a test on Trial 13. Further analyses show that the 16 *S*s in the NI Group having recall on Trial 13 actually had slightly better performance on the first recall-trial than on the last OL trial. The 16 *S*s in the NI Group having immediate recall on Trial 5 showed almost the same loss as did the two subgroups of *S*s in the I Group. These data suggest an interaction between the point of immediate recall and NI vs. I. A 2 × 2 analysis of variance shows the interaction to be significant between the 1% and 5% significance-level. This is true both for loss scores based on projected values and those based on performance on the last trial of OL.

The obvious next question stemming from the above facts is why should this particular group (the group having immediate recall on Trial 13 but not given anti-rehearsal instructions) show the least forgetting of the four groups? Although we cannot attest to the validity of reports from *S*s who say they did *not* systematically rehearse over the retention-interval, positive reports (reports of systematic rehearsal) probably should not be discounted. Of the 16 *S*s in the NI group given recall on Trial 13, 7 said they tried to recall all the items in the list learned at least once during the 24-hr. retention-interval. Of the 16 *S*s in the NI Group given immediate recall on Trial 5, only one gave a report of such extensive rehearsal-activity. Only one of the 32 *S*s given instructions not to rehearse reported such rehearsal. Thus, it would appear that there is a reason for the interaction noted; but the question can be raised as to whether or not the rehearsal facilitated recall. The seven *S*s from the NI Group having immediate recall on Trial 13, and admitting systematic rehearsal, showed a mean gain (using projected scores) of 0.61 items over the interval; the nine *S*s in the same group not admitting rehearsal showed a mean loss of 0.54 items. These data do suggest, then, that rehearsal was positively related to retention. Nevertheless, it is not clear why recall just before the retention interval should tend to make more *S*s rehearse. As will be seen, there is no evidence of this happening when immediate recall was given on the Trial 25. It is, therefore, a distinct possibility that the NI Group given recall on Trial 13 simply contained, by chance, more *S*s who would have rehearsed under any conditions. Finally, it should be noted that relearning over the five trials shows no effect attributable to differences in instructions. For each group the performance on the last five OL trials was subtracted from the performance on the five relearning trials. The difference between these two distributions yielded a *t* of 0.22.

Turning next to the data for the 25-OL Groups, it is seen (Fig. 3) that the effect of instructions not to rehearse is just the opposite of that shown for the 13-OL Groups. After 25 trials the I Group shows less loss than the NI Group. Using performance on the last OL trial as a base, the I Group shows a mean loss of 0.47 items, the NI Group a loss of 1.03 items. The *t* is 2.33. This statistic may not be judged entirely appropriate because of severe skewing of the distributions. A 2 × 2 chi-square table, showing the number of *S*s who showed no loss or a gain and the number who showed a loss, gives, however, a chi-square of 5.08 which is also significant between the 5% and 1% level. There is no evidence of interaction as a function of whether or not *S*s were given a recall on Trial 25.

Among the 32 *S*s of the I Group given 25 OL trials, only one admitted systematic rehearsal; in the NI Group, 8 *S*s admitted the same. The 8 *S*s in the NI Group admitting rehearsal lost a mean of 0.64 items, while the 24 *S*s in this group who did not report rehearsing lost 0.95 items, suggesting that rehearsal did retard forgetting. But over-all, of course, the 32 *S*s in the I Group (with one admitted rehearser) showed less loss than the 32 *S*s in the I Group (with 8 admitted rehearsers). Thus, differences in retention between the two groups do not seem to be attributable to differences in rehearsal. Indeed, there seem to be no apparent reasons why the difference in retention occurred. Of course, it is possible that many *S*s in the NI Group did rehearse but were reluctant to admit this in the post-experimental inquiry. It is also possible that rehearsal is irrelevant to retention. While the data have shown that within a group those admitting rehearsal retained better than those not admitting rehearsal, it is conceivable that those who said they rehearsed would have retained better than the others even if they had not rehearsed. The lack of consistency of the effects of the anti-rehearsal instructions on recall at the two levels of learning makes interpretation difficult. The results at the two levels of learning are consistent in showing that anti-rehearsal instructions had no effect on relearning.

It would seem quite unlikely that control could be exercised over a particular kind of reported 'rehearsal' by any procedure short of keeping *S* occupied in the laboratory. This rehearsal involves items which 'pop in' *S*'s mind for no apparent reason. Figures given earlier on number of *S*s rehearsing included only those who systematically rehearsed—who orally or in writing tried to remember all items. A majority of *S*s in all groups, however, reported the 'pop-in' phenomenon for some items at various times during the 24-hr. interval. It does not seem possible to prevent this phenomenon by instructional procedures.

Over-all retention. An examination of Fig. 3 shows that on an absolute basis retention over the 24-hr. interval was high. If performance on the last trial is used as a base, the retention of the 64 *S*s having 13 OL trials is 91%; for the 64 *S*s having 25 OL trials the value is 90%. The use of projected scores, which would adjust for differences in the acceleration ot the learning curves at the time OL was terminated, changes the value to 87% for the *S*s having 13 OL trials, but does not change that for the *S*s having 25 OL trials.

It will be remembered that the index of associative connection correlated highly with the number of correct anticipations an item received during OL. The mean number of times an item was *recalled* also correlated highly with the index of associative connection. For the 16 items the *r* between the index of associative connections and recall after 13 OL trials is 0.87 and after 25 OL trials, 0.68. However, these relationships probably reflect primarily the differences in number of correct anticipations during OL. The *r* between median number of times an item was correctly anticipated during OL and number of *S*s correctly recalling the item was 0.90 after 13 OL trials, and 0.82 after 25 OL trials.

As noted above, forgetting over 24 hr. was relatively slight, and for the two degrees of learning was essentially a constant proportion of the number of correct anticipations on the last OL trial. Performance on the second relearning trial was about equivalent to the performance on the last OL trial. This very rapid recovery may suggest that even a portion of the approximately 10% loss measured after 24 hr. might be attributed to performance-factors, *e.g.* lack of set or warm-up. In any event, these *S*s, naïve to formal verbal learning, showed very little forgetting. Therefore, an effort was made to get some indication of retention over a much longer interval.

Eleven-month retention. Twenty *S*s were brought back to the laboratory approximately 11 mo. after they had served in the original experiment. Most of these *S*s had served in other experiments during the interval. Therefore, they were carefully instructed concerning this particular experiment, and were asked to try to recall responses on the first trial, guessing if they were not sure. Following this recall-trial, nine 'relearning' trials were given.

Of the 20 *S*s, 9 had had 13 OL trials and 11 had had 25 OL trials. A mean of 11.3 mo. had elapsed since the 13-trial *S*s had originally relearned, and 10.7 mo. for the 25-trial *S*s. As may be seen in Fig. 4, the original learning of these two small groups was quite comparable, and they represent well the original large groups. Of the nine *S*s in the 13-trial group, one *S* recalled four items correctly, and three recalled one each. In these three cases the item recalled was the one with the highest associative connection and might, therefore, be due to guessing. Among the 11 *S*s having had 25 OL trials, one recalled two items, and three recalled one each. Again, the items recalled had high associative connection.

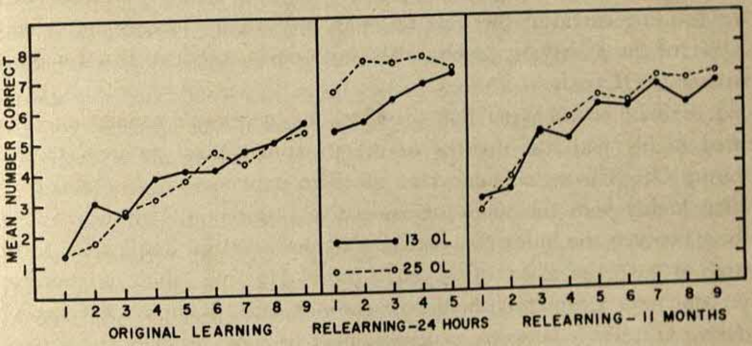


FIG. 4. RELEARNING AFTER 11 MO. COMPARED WITH ORIGINAL LEARNING AND RELEARNING AFTER 24 HR.

In Fig. 4 the first 9 OL trials and the 5 relearning-trials taken 24 hr. after original learning are plotted, followed by the 9 relearning trials taken after 11 mo. The 11-mo. curves indicate more rapid learning than in OL. This may represent a residue of learning-to-learn resulting from this and subsequent experiments in which the *S*s served, or it may represent a residue of memory for the list itself. There is no way to tell the amount contributed by each factor. That the memory for specific items is slight is suggested by the fact that these 11-mo. curves do not differentiate clearly between *S*s having had 13 and 25 OL trials. While the 25-trial *S*s appear to be slightly superior, even this slight superiority might be due to the fact that the interval of retention was 0.6 months shorter for these *S*s than for the 13-trial *S*s. Thus, neither at 24 hr. (where percentage loss was essentially identical for both degrees of OL) nor at 11 mo., is there evidence for a differential *rate* of forgetting as a function of degree of OL.

SUMMARY AND CONCLUSIONS

This study dealt primarily with two problems of method which arise in the study of forgetting of verbal materials. The first problem concerned the base or control measure to use in order that measurements of retention will not be confounded by differences in degree of original learning. This problem arises: (1) whenever retention of different materials is to be compared, and (2) in any situation where comparisons of retention are to be made and where acquisition-curves for the lists to be compared show different rates of acceleration at the point at which the interval of retention is introduced. Carrying learning to the same criterion (*e.g.* one perfect trial) before the retention-interval is introduced does not handle the problem since performance under the various conditions might well have differed on a 'next' trial. If it does, and if not adjusted for, the difference becomes a part of the retention-measurements and is, therefore, a confounding factor.

Three methods of deriving base-measures were compared. In one set of conditions *S*s were given immediate tests of recall at various levels of learning. In another set of conditions *S*s received continuous learning over the same range as included in the immediate-recall conditions. As a third measure, projected scores were used, these scores being based on the past performance of each *S* in the group.

The results have provided firm conclusions concerning these three methods of deriving base-measures. Immediate recall and continuous learning gave essentially the same results. In any particular situation, therefore, it appears that the choice of one of these methods may be made on the basis of convenience or efficiency. The evidence does not recommend the use of projected scores with the low levels of learning. These projected values clearly overestimate performance at low degrees of learning so that absolute measures of forgetting based on them would be seriously in error. The magnitude of this error undoubtedly will be less when lists of homogeneous items are used (as opposed to the heterogeneous lists used here). Nevertheless, since it is probably impossible to construct perfectly homogenous lists, some error in projecting will always be present. At high levels of learning the data indicate that the magnitude of the overestimation will be small even with heterogeneous lists. At such levels of learning, the projected scores seem to be as satisfactory as either the continuous-learning or immediate-recall scores.

The second problem, that of attempting to control rehearsal of *S*s over long intervals, was not solved satisfactorily. Groups of *S*s were given either 13 OL trials or 25 OL trials. Half of the *S*s at each level of learning were given instructions designed to minimize rehearsal over the 24-hr. interval of retention, and half were not given such instructions. After 13 OL trials the instructions not to rehearse might be said to have had a beneficial effect in the sense that the groups so instructed showed more forgetting than those not instructed. A peculiar interaction, however, made this interpretation questionable as a general result. Those *S*s given an immediate test of recall on their last OL trial (Trial 13) and given no instructions about rehearsal, produced most of the variance attributable to instructions. A group given an immediate test on Trial 5 but also not given instructions about rehearsal performed in a manner equivalent to those given instructions. This interaction produced by point of immediate recall is a difficult one to interpret. Moreover, since such a condition which produced the effect (immediate recall on last OL trial) is very unlikely to occur in studies of retention, it would seem that it should be discounted as an important point of method. This conclusion is further emphasized by the results obtained after 25 OL trials. No interaction occurred at this level of learning (although one group was given immediate recall on Trial 25). Furthermore, *S*s given anti-rehearsal instructions showed less forgetting than those not instructed. Thus, the effect was opposite to that found at the lower degree of learning. Indeed, a 2 × 2 analysis of variance of scores of retention with degree of learning and instructions as the two variables showed only the interaction to be significant. None of the evidence examined made this interaction immediately meaningful psychologically. By certain *ad hoc* assumptions (*e.g.* that *S*s instructed not to rehearse in fact rehearse more than those not instructed, and that rehearsal may facilitate or inhibit performance depending upon the level of learning), the basic findings might be accommodated. It must be remembered, however, that none of the effects of the anti-rehearsal instructions on recall attained a high level of statistical significance and no effects were measureable in relearning.

Perhaps other instructional techniques designed to prevent rehearsal would be more effective (and the results more consistent) than those used here. It is sufficient to say that the present experiment has raised some doubts about the use of any instructions intended to prevent rehearsal. In further studies we are undertaking, no attempt is being made to control rehearsal over long intervals of retention.

AVOIDANCE-CONDITIONING IN THE GOLDFISH: EXPORATORY STUDIES OF THE *CS–US* INTERVAL

By ERIKA R. BEHREND and M. E. BITTERMAN, Bryn Mawr College

The experiments to be reported here are concerned with the behavior of the goldfish in a situation patterned after the mammalian shuttle-box. Their purpose is comparative. They are part of a larger program of research designed to delineate similarities and differences in the learning of fish and rat.[1] The principal variable of the present work, which figured prominently also in the earliest work with the mammalian shuttle-box,[2] and in which there has been a recent renewal of interest,[3] is the *CS–US* interval. The use of a delay- rather than a trace-procedure reflects a curiosity about certain mammalian data, to be considered later, which suggest that the *CS–US* interval plays a rather different role than ordinarily is assigned to it in contemporary learning theory.

EXPERIMENT I

In the first experiment, the acquisition of an avoidance-response was studied in six groups of fish, each trained at a different *CS–US* interval. To aid in the interpretation of the results obtained, two additional groups were studied. One group was trained first at a short interval and then shifted to a longer interval, while the reverse procedure was used for a second group.

Method. A detailed description of the apparatus employed is provided elsewhere.[4] There were four shuttle-boxes in which four *S*s were trained simultaneously. Each box was a long, narrow Plexiglas chamber divided into two compartments by a hurdle. Although the water-level in each compartment was several inches in height, the level at the hurdle was less than an inch, which produced some resistance to

---

* Received for publication November 10, 1960. This research was supported by Grant M-2857 from the National Institute of Mental Health.

[1] M. E. Bitterman, Toward a comparative psychology of learning, *Amer. Psychologist,* 15, 1960, 704-712.

[2] L. H. Warner, The association span of the white rat, *J. genet. Psychol.,* 41, 1932, 57-89.

[3] L. J. Kamin, Traumatic avoidance learning: The effects of CS–US interval with a trace-conditioning procedure, *J. comp. physiol. Psychol.,* 47, 1954, 65-72; F. R. Brush, E. S. Brush, and R. L. Solomon, Traumatic avoidance learning: The effects of CS–US interval with a delayed-conditioning procedure, *ibid.,* 48, 1955, 285-293; R. M. Church, F. R. Brush, and R. L. Solomon, Traumatic avoidance learning: The effects of CS–US interval with a delayed-conditioning procedure in a free-responding situation, *ibid.,* 49, 1956, 301-308.

[4] J. L. Horner, Nicholas Longo, and M. E. Bitterman, A shuttle box for the fish and a control circuit of general applicability, this JOURNAL, 74, 1961, 114-120.

crossing (and so tended to give the response an all-or-none character), but the resistance was not so great that it could not be overcome easily by mild shock. Interval-responses were permitted, and they occurred with substantial frequency.

Movement from one compartment to another was detected by means of photocells, and the position of each *S* in each box was monitored continuously by ratchet-relays. Trials were programmed by *E,* who operated a selector-switch to determine which animal in the set of four was to have a trial; pressed a button to start the trial; and, when the trial was over, read the latency of response (to the nearest 0.01 sec.) from a Standard Electric timer, which started with the onset of the *CS* and was stopped by *S*'s response. The interval-responses made by each *S* were registered on a counter, which was read by *E* at the conclusion of each series of trials.

Each trial began with the onset of light in the compartment occupied by *S,* and the light was followed after a preset interval (the *CS–US* interval) by intermittent shock (a 0.25-sec. pulse about every 1.5 sec.). Intermittent shock was used because continuous shock intense enough to be consistently effective tended to produce tetany. The level of shock used, which was established in pretests with a group of *S*s not used in the experiment, was just high enough to ensure prompt and consistent escape. Each trial was terminated either by the response of the animal (changing compartments) or, independently of response, 15 sec. after onset of the *US*. (Only very rarely, however, did an animal fail to escape shock in the 15-sec. period.) Response to light alone terminated the light and avoided shock. Response to light and shock terminated both stimuli.

The *S*s were 80 goldfish, 2–3 in. in length, obtained from a local dealer. They were divided at random into eight groups of 10 *S*s each. Six groups were trained for 20 days with intervals of 2.5, 5, 10, 20, 40, and 60 sec., respectively. A seventh group was trained for the first 10 days with a 2.5-sec. interval and for the next 10 days with a 20-sec. interval. An eighth group was trained for the first 10 days with a 20-sec. interval and for the next 10 days with a 2.5-sec. interval. Each animal was given 10 trials per day, at about the same time each day, on six days of each calendar week. The mean intertrial interval was about 3 min., ranging approximately from 2–4 min. The *S*s lived in individual 2-gal. tanks on open shelves in a temperature-controlled laboratory. They were transferred to and from the apparatus by means of a small net.

Results. The course of learning in the six groups trained with different *CS–US* intervals is plotted in Fig. 1 in terms of mean number of avoidances per day. The curves are negatively accelerated in shape and seem to reach different asymptotes (about 80% for the 2.5- and 5-sec. groups, and about 95% for the others). In Fig. 2, rate of learning as a function of *CS–US* interval is plotted in terms of mean trials to three different criteria: first avoidance, 9 avoidances in the 10 trials of any day, and 10 avoidances on any day. No point is plotted for the most stringent criterion (10/10) at the shortest interval (2.5 sec.), because not all of the animals in the group (actually very few of them) ever had a perfect day. The three curves (and others which may be plotted for still other criteria) suggest that the difficulty of the problem declines in negatively accelerated fashion as the *CS–US* interval increases, reaching asymptote somewhere between 20 and 40 sec. A simple analysis of variance based on trials to the criterion of intermediate stringency shows the effect of *CS–US* interval to be highly reliable ($F = 10.09$ with 5 and 54 *df.*, $p < 0.01$).

In his early study of rats in a shuttle-box, Warner found a rather sharp *increase* in trials to criterion as the *CS–US* interval was lengthened (from 1–30 sec.), and so also did Kamin in a more recent study of dogs (5–40 sec.).[5] In both these cases, a trace-conditioning procedure was used; the *CS* (auditory) was sounded only briefly (for 1 sec. in the early experiment and for 2 sec. in the later one), and the interval between its offset and the onset of shock was silent. It seems, however, that the trace-procedure and the delay-procedure give different results. Church, Brush, and Solomon, working with dogs under conditions much like Kamin's, but with a delay-procedure, found no significant relation between rate of learning and *CS–US* interval over the range of 5–20 sec.[6] Another delayed-conditioning experi-

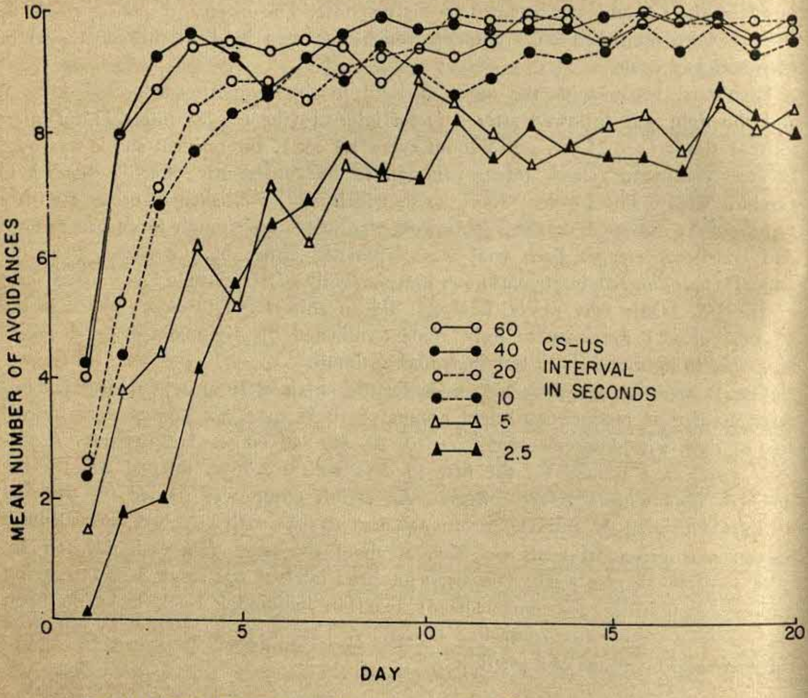


FIG. 1. MEAN NUMBER OF AVOIDANCES PER DAY AS A FUNCTION OF *CS–US* INTERVAL (EXPERIMENT I)

ment with dogs (under somewhat different conditions) by Brush, Brush, and Solomon also failed to yield marked or consistent relations between various measures of the rate of learning and *CS–US* interval (2.5–80 sec.).[7]

---

[5] Warner, *op. cit.*, 79; Kamin, *op. cit.*, 67.

[6] Church, Brush, and Solomon, *op. cit.*, 303.

A clear inverse relation over a wide range of *CS–US* intervals between trials to criterion and interval, such as that which appears in the present data for the goldfish, has not heretofore been reported, although subsequent work with another fish has yielded much the same function.[8] We come here upon the only suggestion of a species-difference which turned up in this experiment. Whether a species-difference is indeed involved, however, it is, of course, too early to say. The difference in outcome may perhaps be attributable to differences in such variables as shock-intensity. In the experiments with dogs, the greatest intensity feasible was used; in the present experiment, the lowest which would produce consistent escape.

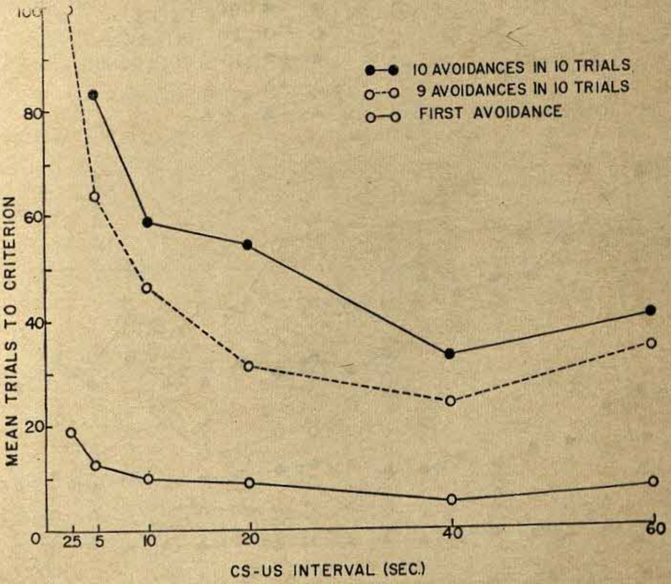


FIG. 2. TRIALS TO CRITERION AS A FUNCTION OF *CS–US* INTERVAL (EXPERIMENT I)

In Fig. 3, change in avoidance-latency over the 20 days of training is shown for each of the six groups. (These and all other such values presented in this paper are uncontaminated by failures to avoid; that is, they

---

[7] Brush, Brush, and Solomon, *op. cit.*, 286.

[8] Jerome Wodinsky, E. R. Behrend, and M. E. Bitterman, Avoidance-conditioning in two species of fish, *Anim. Behav.*, 10, 1962, in press.

27-8-76
LIBRARY
9218

are based only on actual avoidances.) The latencies decline in negatively accelerated fashion to different asymptotes which vary directly with *CS–US* interval. The relation between mean avoidance-latency and *CS–US* interval at different stages of training (the first day on which there were 9 avoidances, the first day on which there were 10 avoidances, and the twentieth day of training) is shown in Fig. 4. The plots are essentially linear. There is a decline in slope as training progresses, but, even on the twentieth day —long after the achievement of an asymptotic level of performance— there is a marked tendency for latency to increase with *CS–US* interval which is statistically reliable ($F = 19.07$ with 5 and 54 *df*, $p < 0.01$). Experiments with dogs have yielded the same result, both for the trace- and for the delay-procedure.[9] The absolute avoidance-latencies for the various *CS–US* intervals recorded in the present experiment actually were very

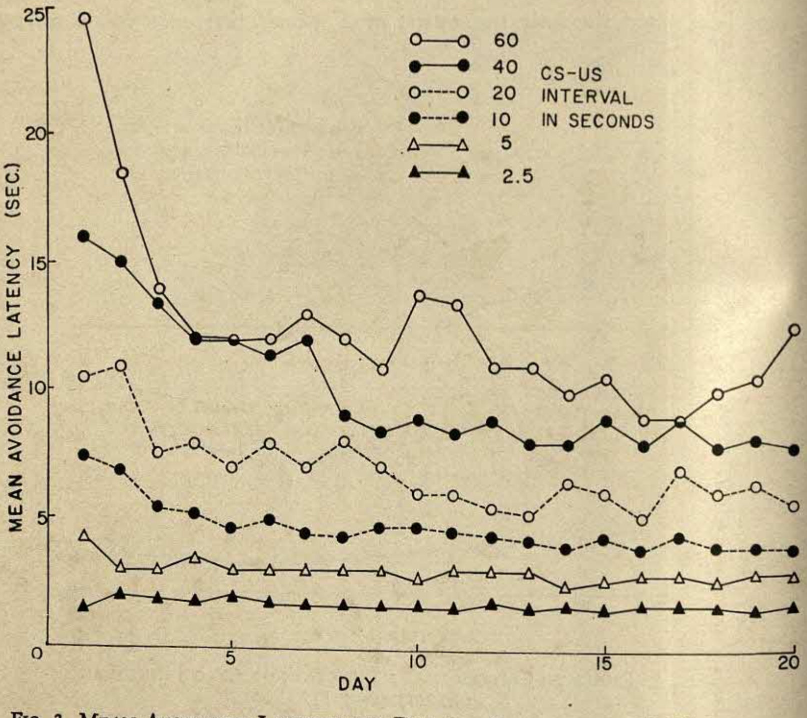


FIG. 3. MEAN AVOIDANCE-LATENCY PER DAY AS A FUNCTION OF *CS–US* INTERVAL (EXPERIMENT I)

---

[9] Kamin, *op. cit.*, 68; Brush, Brush, and Solomon, *op. cit.*, 288; Church, Brush, and Solomon, *op. cit.*, 304.

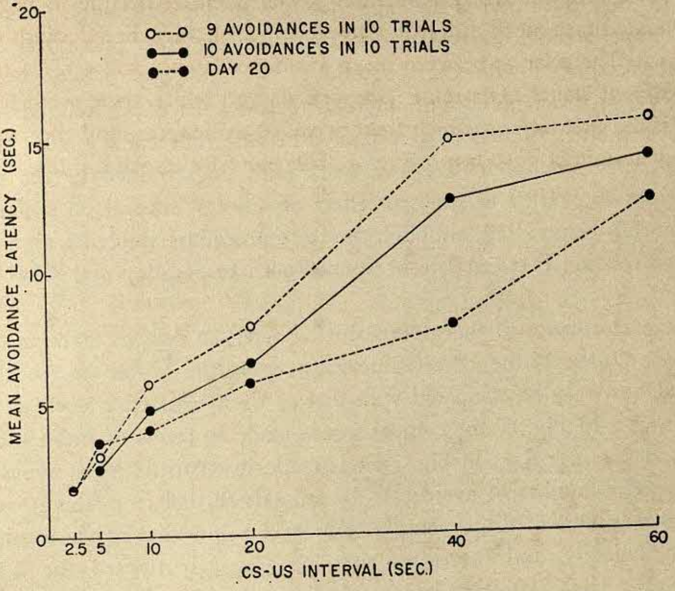


FIG. 4. MEAN AVOIDANCE-LATENCY AT VARIOUS STAGES OF TRAINING AS A FUNCTION OF *CS–US* INTERVAL (EXPERIMENT I)

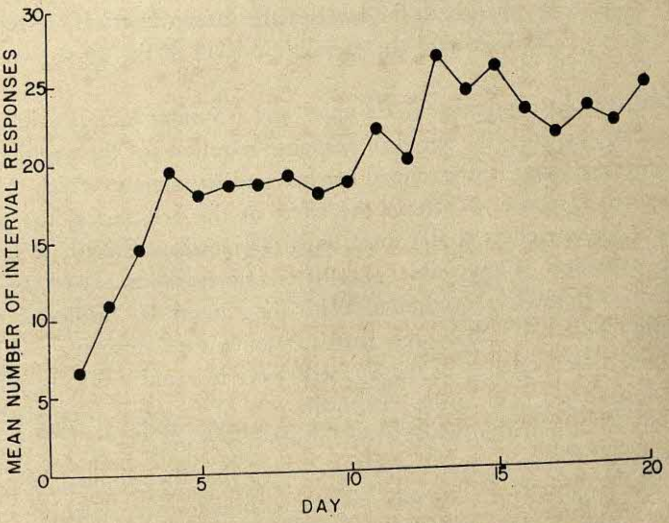


FIG. 5. MEAN NUMBER OF INTERVAL-RESPONSES PER DAY (EXPERIMENT I)

close to those measured in the experiments with dogs, although the asymptotic escape-latencies were somewhat lower (on the order of a fraction of a second as compared with several seconds).

In Fig. 5, mean number of interval-responses per day is plotted for all groups combined. The data were pooled because differences among groups proved to be negligible. Previous work with dogs trained by the delay-procedure also failed to show an effect of *CS–US* interval on amount of spontaneous responding, although the trace-procedure yielded a significant positive relation between rate of spontaneous responding and *CS–US* interval.[10]

The performance of the groups shifted after 10 days of training either from the 20-sec. to the 2.5-sec. interval or from the 2.5-sec. to the 20-sec. interval now may be compared with that of the unshifted 2.5-sec. and 20-sec. groups. In Fig. 6, the comparison is made in terms of mean number of avoidances per day; in Fig. 7, it is made in terms of mean avoidance-latency. The number of avoidances by animals shifted from the 20-sec. to the 2.5-sec. interval drops sharply to a level below that of the unshifted 2.5-sec. animals, and thereafter rises progressively toward the 2.5-sec. asymptote. That avoidance-latency declines sharply after the shift is a necessary consequence of the fact that only responses of short latency would avoid shock. The number of avoidances made by animals shifted from 2.5-sec. to 20-sec. rises sharply to the level of the 20-sec. controls. Avoidance-latency rises gradually toward the asymptotic level of the 20-sec. controls.

Discussion. As contiguity of *CS*-onset and *US*-onset increases in an experiment of this sort, the speed of response required for the avoidance of shock increases; that is, the animal has less and less time in which to make the required response. Whatever the effect of the first factor, the second (where it has any effect at all) must impair performance. With an interval of zero, avoidance is impossible, and there is no reason to expect the function to pass through a maximum. With long intervals, furthermore, the probability becomes rather high that a 'spontaneous' response, which is made quite independently of the *CS*, will serve to avoid shock. Insufficient attention has been paid in the past to the role of factors confounded with contiguity of stimulation which may improve performance in experiments on avoidance. Solomon and Brush have noted that the minimal latency of a response required for avoidance may be such as to make learning impossible when the *CS–US* interval is very brief, but the data available to them

---

[10] Kamin, *op. cit.*, 69; Church, Brush, and Solomon, *op. cit.*, 305.

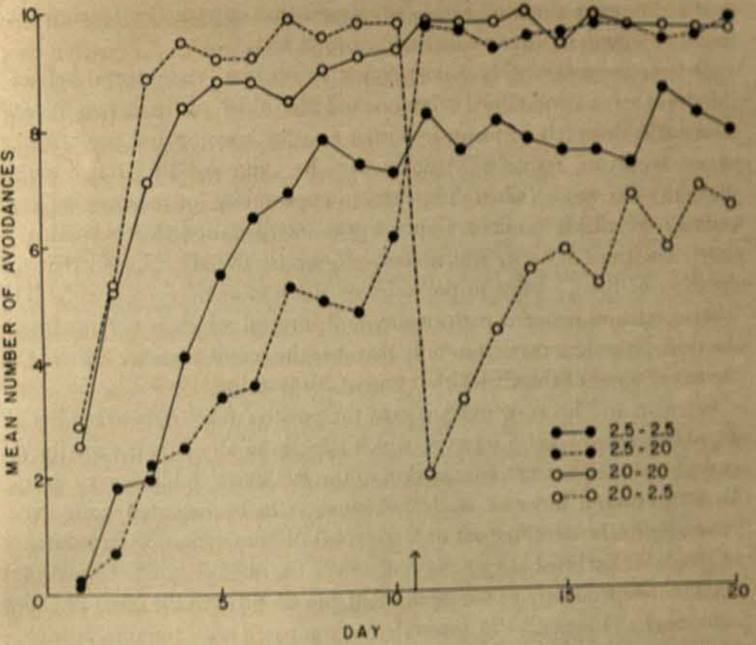


FIG. 6. MEAN NUMBER OF AVOIDANCES PER DAY IN THE TWO SHIFTED GROUPS AND THE CONTROLS (EXPERIMENT I)

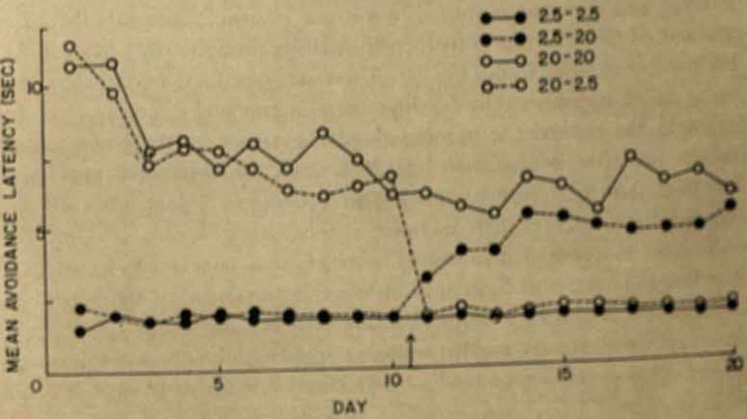


FIG. 7. MEAN AVOIDANCE-LATENCY PER DAY IN THE TWO SHIFTED GROUPS AND THE CONTROLS (EXPERIMENT I)

at the time gave them no reason to suspect that opportunity for response might be important over a wider range of intervals.[11]

Despite the assertion by Solomon and Brush that "the interval between the *onset* of a conditioned stimulus and the *onset* of an unconditioned stimulus is definitely of major importance in the learning and performance of an avoidance response,"[12] doubt may be expressed that this variable plays any role at all (when differences in opportunity for response are controlled). Available results are open to the interpretation that it is not *CS*-onset, but the *CS* itself, which comes to arouse the animal, and that the duration of the *CS* prior to the onset of shock is of no consequence. The inverse relation between performance and interval which is obtained with the trace-procedure may mean only that it is the recent *presence* rather than the recent *onset* of the *CS* which is crucial for learning.

Solomon and his co-workers regard the positive relation between avoidance-latency and *CS–US* interval, which appears in all of their experiments as well as in the present one, as akin to the Pavlovian 'inhibition of delay.' To some extent, however, such a relation is to be expected quite independently of the development of a temporal discrimination. Even if latency of avoidance declined as a simple function of the number of *CS–US* pairings —quite independently of the duration of the *CS* prior to the onset of shock —training at lower *CS–US* intervals (as compared with training at higher *CS–US* intervals) would produce lower avoidance-latencies, because only lower-latency responses would be counted as avoidances. The effect can be seen in Fig. 7. Two groups of animals are trained at the same (20-sec.) interval, and then one is shifted to a lower (2.5-sec.) interval; the mean avoidance-latency of the shifted group declines precipitously—as does also the number of avoidances (Fig. 6). There are some data, nevertheless, that point unquestionably to the development of a temporal discrimination: for example, the gradual rise in mean avoidance-latency for the group trained at 2.5 sec. after it was shifted to 20 sec., or the differential avoidance-latencies shown by four of the groups (10–60 sec.) long after all had reached the same very high asymptotic probability of avoidance. Until disinhibition has been demonstrated, however, it is unnecessary to attribute the longer latencies of *S*s trained at longer intervals to the differential inhibition of tendencies to respond in the early seconds of the *CS*. It seems more reasonable to assume that animals trained with longer intervals simply never develop very strong tendencies to respond in the early seconds of the

---

[11] M. R. Jones, ed., *Nebraska Symposium on Motivation,* 1956, 260-261; Marvin Schwartz (Conditioned-stimulus variables in avoidance learning, *J. exp. Psychol.,* 55, 1958, 347-351), finding better performance at 6 sec. as compared with 3 sec., subsequently remarked on the same problem.

[12] Jones, *op. cit.,* 264.

CS. This interpretation is supported by the negative transfer found for the group trained at 20 sec. and shifted to 2.5 sec. when its performance is compared with that of the group trained from the outset at 2.5 sec. (Fig. 6).

As has been noted by Brush, Brush, and Solomon, the attribution of noxious properties to the *CS* and of reinforcing properties to its termination leads to the prediction that the avoidance-latencies of all groups should approach the same low value despite differences in *CS–US* interval.[13] The contrary results which appear consistently in these experiments suggest the possibility that the function of the *CS* may be entirely discriminative. The fact that an animal will learn to make a response that terminates a stimulus which has been paired with shock does not mean necessarily that the stimulus has acquired noxious properties; in an experiment by Sidman and Boren, rats responded at a high level when response postponed a signal followed by unavoidable shock, but not when response postponed a signal for avoidable shock.[14] Nor does the fact that a stimulus which has been paired with food will, under certain circumstances, increase the probability of a response which produces it necessarily imply that the stimulus has acquired rewarding properties; when the discriminative properties of such a stimulus are pitted against whatever rewarding properties it may have, the former are found to dominate.[15] The oversimplified doctrines of acquired drive and secondary reinforcement are badly in need of reëvaluation.

EXPERIMENT II

The purpose of this experiment was to study the effect of *CS–US* interval in training on resistance to extinction. The design was factorial: there were three *CS–US* intervals in training and three *CS*-durations in extinction, yielding a total of nine different conditions.

In previous experiments, the practice has been to use but a single *CS*-duration in extinction, and the role of this variable has been systematically underestimated. For example, in speculating about the lesser resistance to extinction found in their own experiment as compared with an earlier one, Church, Brush, and Solomon considered a variety of possibilities except the duration of the *CS* in extinction, although in their experiment *S* had 30 sec. in which to respond, while in the earlier experiment 2 min. were allowed.[16] A factorial design makes it possible not only to evaluate the effect of *CS*-duration in extinction but to discover any interaction which may exist between *CS–US* interval in training and *CS*-duration in extinction.

Method. The apparatus and general procedure were exactly the same as in Experiment I. The *S*s were 90 goldfish, of the same type and drawn from the same population as those of the earlier experiment. They were divided at random into three groups, of 30 animals each, which were trained with *CS–US* intervals of 5, 10, and 20 sec., respectively. There were 10 trials per day with a mean intertrial interval of 3 min., and the criterion of learning was 10 avoidances on a single day. On the day following that on which each animal attained the criterion, its extinction was begun. Ten of the animals in each major group were extinguished with a 5-sec. *CS*-duration. (That is, any response which occurred less than 5 sec. after the onset of the *CS* terminated the *CS* and was counted as an avoidance, but, if the animal failed to respond within 5 sec., the *CS* terminated automatically and a failure of avoidance was recorded.) Ten of the animals in each group were extinguished with a *CS*-duration of 10 sec., while the *CS*-duration for the remaining third of each group was 20 sec. The criterion of extinction was two or fewer avoidances on a single day, but work with an animal was terminated in any case if it had failed to reach the criterion on the fifteenth day of extinction. Since the various animals in each of the main training-groups reached criterion at different times, it was not possible to have strictly matched extinction-subgroups, but some rough matching could be attempted, and the outcome actually was quite satisfactory.

---

[13] Brush, Brush, and Solomon, *op. cit.*, 292.

[14] Murray Sidman and J. J. Boren, A comparison of two types of warning stimulus in an avoidance situation, *J. comp. physiol. Psychol.*, 50, 1957, 282-287.

[15] C. B. Elam, D. W. Tyler, and M. E. Bitterman, A further study of secondary reinforcement and the discrimination hypothesis, *J. comp. physiol. Psychol.*, 47, 1954, 381-384.

[16] Church, Brush, and Solomon, *op. cit.*, 307.

Results. The course of learning in the three main groups was very much the same as that for the groups trained at the same intervals in Experiment I. Again there was a significant inverse relation between *CS–US* interval and trials to criterion, and again a significant positive relation between *CS–US* interval and avoidance-latency on the criterion day. This close correspondance between the results of the two experiments indicates the reliability of the method.

The data on extinction are summarized in Fig. 8, where the relation between mean number of responses in extinction and *CS*-duration in extinction is plotted separately for each *CS–US* interval in training. The curves suggest that resistance to extinction is positively related to *CS*-duration in extinction (more responses in extinction with longer durations) and inversely related to *CS–US* interval in training (more responses in extinction with shorter training-intervals). This impression is supported by the results of an analysis of variance. The effect of *CS*-duration in extinction is significant beyond the 1% level ($F = 43.90$ with 2 and 81 *df.*) as is the effect of *CS–US* interval in training ($F = 8.48$ with 2 and 81 *df.*), while the interaction is negligible ($F < 1$ with 4 and 81 *df.*). In neither of Solomon's experiments with the delay-procedure did a consistent relation appear between *CS–US* interval in training and resistance to extinction.[17] Here again the possibility of a species-difference arises, but again the possibility must be considered that some parametric difference, such as in shock-intensity, was responsible. In Kamin's experiment with the trace-procedure, as in the present experiment, extinction was inversely related to *CS–US* interval in training.[18]

---

[17] Brush, Brush, and Solomon, *op. cit.*, 288-289; Church, Brush, and Solomon, *op. cit.*, 304. An experiment with rats by G. B. Kish (Avoidance learning to the onset and cessation of conditioned stimulus energy, *J. exp. Psychol.*, 50, 1955, 31-38) shows greater resistance to extinction after training with a *CS–US* interval of 0.25 sec. as compared with 1 sec., but no greater resistance after training at 5 sec. than at 1 sec.

In the course of extinction, avoidance-latencies seemed to increase progressively in all groups. The change could best be charted in the groups extinguished at 20 sec., whose probabilities of avoidance remained high throughout the 15 days of extinction. For the animals trained at 5 sec. and extinguished at 20 sec., the mean avoidance-latency was 4.6 sec. on the first day of extinction, 6.7 sec. on the eighth day, and 7.7 sec. on the fifteenth day. (None of these *S*s met the criterion of extinction.) For the 9 *S*s which failed to extinguish in the group trained at 10 sec. and extinguished at 20 sec., the corresponding values were 5.4, 6.7, and 7.8 sec. For the 7 *S*s which did not extinguish in the group trained and extinguished at 20 sec., the values were 7.2, 9.5, and 9.2 sec. By Friedman's nonparametric test, the progressive increase in latency is significant at a level of confidence well beyond 1%.

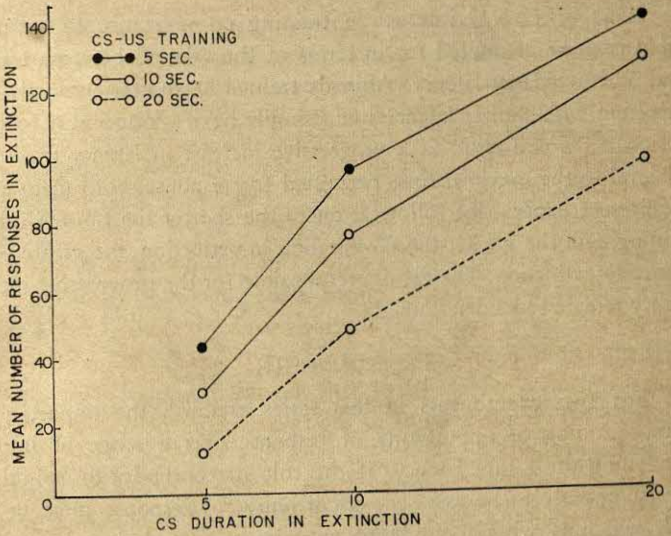


FIG. 8. MEAN NUMBER OF RESPONSES IN EXTINCTION AS A FUNCTION OF *CS*-DURATION AND OF *CS–US* INTERVAL IN TRAINING (EXPERIMENT II)

Discussion. These results emphasize the importance of opportunity for response in the measurement of resistance to extinction. It is obvious that apparent resistance to extinction must be zero when the time permitted for response is less than the minimal time necessary for response, but the present data show that *CS*-duration is important over a range of values greater by a factor of 20 than the minimal response-latency. Whether resistance would continue to improve with further increases in *CS*-duration cannot be said, but it is interesting to note that in the experiments with dogs which yielded extremely high resistance to extinction the *CS*-duration was 2 min.[19]

---

[18] Kamin, *op. cit.*, 67-68.

The effect of *CS–US* interval in training on resistance to extinction probably is to be accounted for in terms of the relation between training interval and avoidance-latency. Animals trained at shorter intervals come to extinction with shorter latencies and simply have a longer way to go. If extinction is the end-result of a progressive increase in latency from some initial level to the maximal time permitted for response, both main effects of the present experiment follow at once; the shorter the *CS–US* interval in training, and the greater the *CS*-duration in extinction, the greater is the resistance to extinction. The factors responsible for the progressive increase in latency remain to be determined.

EXPERIMENT III

The first two experiments in this series establish the importance in avoidance-training of opportunity of response over a range of intervals far exceeding the escape-latency. If the role of contiguity of stimulation *per se* is properly to be assessed, opportunity for response must be controlled, which is what the present experiment was designed to do. A transfer method was employed. Independent groups were given small amounts of training at a variety of different intervals; then all were shifted to the same long interval and their performances compared.

Method. The *S*s were 324 goldfish drawn from the same population as those of the earlier experiments. They were divided at random into 30 groups and trained in three sets. Set 1 consisted of 3 groups of 10 *S*s each; Set 2 consisted of 15 groups of 10 *S*s each; and Set 3 consisted of 12 groups of 12 *S*s each. The apparatus was that of the previous experiments, and as before there were 10 trials per day with a mean intertrial interval of 3 min.

The groups of Set 1 were trained with a *CS–US* interval of zero. That is, light-onset and shock-onset were simultaneous, and avoidance consequently impossible. One group had one day of this simultaneous training, the second had two days, and the third four days, following which all were shifted to a 20-sec. interval for one day (with failure to avoid being punished by shock as before).

---

[19] R. L. Solomon, L. J. Kamin, and L. C. Wynne, Traumatic avoidance learning: The outcomes of several extinction procedures with dogs, *J. abnorm. soc. Psychol.*, 48, 1953, 292; Brush, Brush, and Solomon, *op. cit.*, 286.

The groups of Set 2 were given either one, two, or four days of training with a *CS–US* interval of either 1.25, 2.5, 5, 10, or 20 sec., following which each group was given one day of training with a 20-sec. interval. The groups of Set 3 were given one, two, or four days of training with a *CS–US* interval of 7.5, 15, 30, or 60 sec., following which each group was given one more day of training at a 20-sec. interval. For all sets combined, then, there was a factorial design involving three degrees of training (one, two, or four days) and 10 *CS–US* intervals (ranging from 0–60 sec.). The test in each case was one day of training with a 20-sec. interval.

Results. Mean number of avoidances on the test day is plotted for all groups in Fig. 9. The effect of number of days of training shows up clearly in the curves; except at the zero-interval, performance is best after four days and poorest after one day. The fact that different amounts of training do not produce significant differences in the performance of the three simultaneous groups suggests that their behavior may be due simply to sensitization. This interpretation is supported by results obtained subsequently with an independent group of 10 *Ss* trained for four days with no *CS* at all (a pure escape-condition) and then tested for one day with a 20-sec. *CS–US* interval. The performance of this group actually was better (though not significantly so) than that of the simultaneous groups.

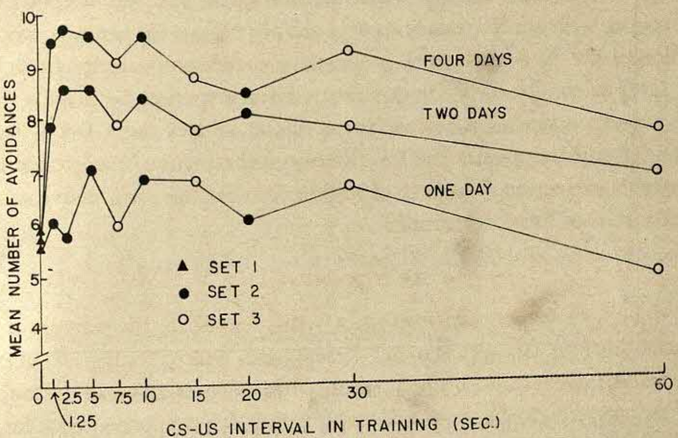


FIG. 9. MEAN NUMBER OF AVOIDANCES ON THE TEST-DAY AS A FUNCTION OF AMOUNT OF TRAINING AND *CS–US* INTERVAL (EXPERIMENT III)

As the *CS–US* interval is increased beyond zero, the number of avoidances rises sharply and then seems to fall off gradually. The decline is not, however, statistically reliable. An analysis of variance for the combined data of Sets B and C—that is, for all intervals except zero—shows a significant effect of number of days of training ($F = 39.88$ with 2 and 268 *df*, $p < 0.01$) but not of *CS–US* interval ($F = 1.22$, with 8 and 268 *df*).[20] The interaction is negligible.

Mean avoidance-latencies are plotted in Fig. 10. Again in terms of this measure, the three simultaneous groups are indistinguishable. For all other intervals except 20 sec., latency declines systematically with number of days of training. The marked inversion at 20 sec., is due, clearly, to error; the performance of other fish trained at 20 sec. in other experiments suggests that the four-day point should be more like 7.5 sec. It also may be seen from Fig. 10 that latency increases as the *CS–US* interval is increased from 1.25–60 sec. Analysis of variance for the data of Sets 2 and 3 combined shows the effect of number of training days to be significant beyond the 5% level ($F = 3.56$ with 2 and 268 *df.*) and *CS–US* interval to be significant beyond the 1% level ($F = 13.53$ with 8 and 268 *df.*). The interaction, too, is significant ($F = 2.17$ with 16 and 268 *df.*, $p < 0.01$). The function for latency suggests that the use of a substantially smaller *CS–US* interval in training would have yielded a very different function for number of avoidances—a function declining sharply with increases in the training interval beyond a few seconds.

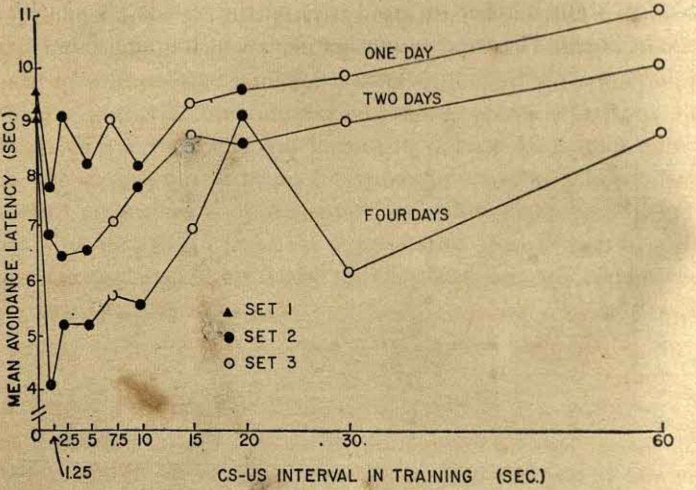


FIG. 10. MEAN AVOIDANCE-LATENCY ON THE TEST-DAY AS A FUNCTION OF AMOUNT OF TRAINING AND *CS–US* INTERVAL (EXPERIMENT III)

It should be noted that the mean number of shocks received in training declined systematically as the *CS–US* interval increased. For the four 10-day groups, the means were 40 (at 0 sec.), 39 (at 1.25 sec.), 34, 25, 22, 19, 11, 16, 10, and 10 (at 60 sec.). It was despite this biasing factor that probability of avoidance did not fall reliably as the training interval increased. The substantial relation between mean avoidance-latency and training interval may, of course, have been due in large measure to this factor.

---

[20] As the degrees of freedom indicate, three *Ss* were lost in the course of the experiment.

Discussion. This experiment demonstrates that the *CS–US* interval plays an important role in avoidance-conditioning which is quite independent of opportunity for response. Strictly simultaneous pairing of the two stimuli apparently produces no true conditioning, a finding which supports the opinion of Pavlov that some precedence of *US* by the *CS* is essential.[21] Just how *much* precedence is essential, the present data do not tell, but a subsequent pilot study provides some indication that conditioning will take place at an interval of 0.62 sec. A group of 10 *S*s trained at that interval for two days (during which time there was not a single avoidance) made a mean of about 9 avoidances in 10 trials when tested at 20 sec. on the third day. A careful mapping of the 0–1 sec. range seems to be in order.

The results for intervals greater than zero contradict the common textbook-generalization (which is based primarily on experiments with human *S*s) that the effectiveness of conditioning declines sharply as the *CS–US* interval is increased beyond a brief optimal value (often given as 0.5 sec.). There is some suggestion in the curves of Fig. 9 of a decline in performance at 60 sec., but the decline is small and statistically unreliable. Interestingly enough, there are other animal data, long neglected, which suggest that rather stable conditioning may be achieved at quite extended *CS–US* intervals. Estes and Skinner, working with rats, got quite good results with an interval of 3 min.; pairing light and shock (delay-procedure), they used as their measure of conditioning the depressive effect of the light on the rate of food-rewarded fixed-interval responding.[22] A similar method was used about 10 yr. ago by Libby in another study of the *CS–US* interval which unconfounds contiguity of stimulation and opportunity for response.[23] Libby trained seven groups of rats by the delay-procedure with intervals between the onset of light and the onset of a 1-sec. shock ranging from 0–30 sec. (30 pairings for each group), and then tested all groups under identical conditions (comparing rates of response during 10-min. periods in light and in dark). The effectiveness of the

---

[21] I. P. Pavlov, *Conditioned Reflexes,* 27; See also Jerome Cohen, Observations on strictly simultaneous conditioned reflexes, *J. comp. physiol. Psychol.,* 43, 1950, 211-216.

[22] W. K. Estes and B. F. Skinner, Some quantitative properties of anxiety, *J. exp. Psychol.,* 29, 1941, 390-400.

[23] Alfred Libby, Two variables in the acquisition of depressant properties by a stimulus, *J. exp. Psychol.,* 41, 1951, 100-107.

CS increased from 0–7 sec. (a less abrupt climb than that found here), but then remained essentially constant with further increases in the interval.

Different results were obtained in a recent study of another species of fish.[24] Six independent groups were conditioned by the delay-procedure with *CS–US* intervals of 0.5–4.0 sec. The *CS* was an increment in illumination and the *US* was a 1-sec. shock. The index of conditioning used—percentage of responses to the *CS* in 60 trials—increased from a low at 0.5 sec. to a maximum at 2 sec. and then declined as the *CS–US* interval was further increased. Opportunity for response was not controlled, but the authors report that responses with latencies of more than 0.5 sec. occurred in fewer than 5% of the trials. In any case, of course, the *decline* in the measure of conditioning at intervals greater than 2 sec. cannot be accounted for in terms of *increasing* opportunity for response. Clearly, much work remains to be done before generalizations about the role of contiguity of stimulation in conditioning can be made with any confidence. The effects of differences in procedure, in measure, and in species must be sorted out carefully in subsequent experiments.

SUMMARY

Avoidance-conditioning in the goldfish was studied in an apparatus patterned after the mammalian shuttle-box. Experiment I was concerned with the role of the *CS–US* interval in acquisition. As the interval increased, there was a negatively accelerated decline in trials to criterion and an approximately linear increase in asymptotic avoidance-latency. In Experiment II, the effect of *CS–US* interval on resistance to extinction was studied. Resistance was found to vary inversely with *CS–US* interval in training and directly with *CS*-duration (opportunity for response) in extinction. In Experiment III, a transfer-design was employed to study the role of the *CS–US* interval in acquisition independently of opportunity for response. A zero-interval (strict simultaneity) produced no conditioning (only sensitization). Over a wide range of positive intervals (1.25–60 sec.), acquisition did not vary significantly, although avoidance-latency in testing increased with the training interval. The results are compared with those of previous experiments on higher animals.

---

[24] Merrill Noble, Anne Gruender, and D. R. Meyer, Conditioning in fish (*Mollienisi sp.*) as a function of the interval between CS and US, *J. comp. physiol. Psychol.*, 52, 1959, 236-239.

CHOICES AMONG UNCERTAIN OUTCOMES: A TEST OF A DECOMPOSITION AND TWO ASSUMPTIONS OF TRANSITIVITY

By BETTY J. GRISWOLD, Columbia University, and
R. DUNCAN LUCE, University of Pennsylvania

During the past few years, considerable attention has been devoted to mathematical descriptions of choices among uncertain outcomes (gambles) and the theories based on these descriptions have led to numerous empirical studies.[1] The guiding principle has been that choices can be represented as a maximization of expected utility. More exactly, the idea is to assign, for every subject (*S*), utilities to gambles and subjective probabilities to events so that (1) the utility of a gamble equals the expectation of the utilities of the constituent outcomes of the gamble relative to the subjective probabilities of the events upon which these outcomes are contingent; and (2) the ordering of the gambles by utility reflects *S*'s preferences. The earliest theories were not probabilistic, which made them difficult to test because, even in a short series of judgments, the *S*s often make intransitive and inconsistent choices. Whether this is due to some change in *S* during the experiment or whether it indicates a probabilistic determination of choices that permits certain local inconsistencies is not easy to decide. The second alternative is, however, simpler to formalize, hence it is explored first.

Our starting point is a probabilistic theory that has been presented in

---

* Received for publication November 18, 1958. This experiment was conducted at the Bureau of Applied Social Research, Columbia University, and was supported in part by National Science Foundation grants NSF-G2803 and G4506 and by National Institute of Health grant M-2293. Our thanks are due Professors Eugene Galanter and William J. McGill for advice concerning the design of this study.

[1] See J. S. Chipman, Stochastic choice and subjective probability, in D. Willner (ed.), *Decisions, Values, and Groups,* 1960, 70-95. C. H. Coombs, On the use of inconsistency of preferences in psychological measurement, *J. exp. Psychol.,* 55, 1958, 1-7; Donald Davidson and J. Marschak, Experimental tests of a stochastic decision theory, in C. W. Churchman and Philburn Ratoosh (eds.) *Measurement: Definitions and Theories,* 1959, 233-269; Donald Davidson, Patrick Suppes, and Sidney Siegel, *Decision Making,* 1957; and Frederick Mosteller and Philip Nogee, An experimental measurement of utility, *J. pol. Econ.,* 59, 1951, 371-404. Two general surveys of work in this area are contained in Ward Edwards, The theory of decision making, *Psychol. Bull.,* 51, 1954, 380-417, and R. D. Luce and Howard Raiffa, *Games and Decisions,* 1957.

the economic literature.[2] We will not attempt a detailed summary here, but several features must be mentioned.

First, an explanation of the notation: Lower case Latin letters 'a,' 'b,' . . . . denote pure outcomes, such as receiving a book or losing a sum of money, and lower case Greek letters 'α,' 'β,' . . . denote events whose occurrences are uncertain, such as "a six coming up on a throw of a die" or "it will rain in Boston tomorrow." A symbol of the form 'aαb' denotes the gamble in which 'a' is the outcome if the event α occurs and 'b,' if it does not. It is assumed that if two gambles, 'aαb' and 'cβd,' are presented to *S*, then there is a probability $P(a\alpha b, c\beta d)$ that *S* will choose 'aαb' in preference to 'cβd.' Although it is not clear how to introduce expected utility into a probabilistic framework, the intuition that choices between two gambles can be decomposed by considering their constituent parts—outcomes and events—can be retained in a much weakened form.

Suppose that there is a probability $Q(\alpha,\beta)$ that *S* judges Event α more likely to occur than Event β. Moreover, suppose that *S*'s preferences between pairs of outcomes are statistically independent of his judgments about the likelihood of events, then it is not difficult to see that

$$P(a\alpha b, a\beta b) = P(a, b)Q(\alpha, \beta) + P(b, a)Q(\beta, \alpha) \cdot \cdot \cdot \cdot \qquad [1]$$

We take Equation [1], known as the *decomposition-axiom,* as a basic assumption to be tested. The easiest way to do so is to note that it is equivalent to

$$Q(\alpha, \beta) = \frac{P(a\alpha b, a\beta b)P(a, b) - P(a\beta b, a\alpha b)P(b, a)}{P(a, b)^2 - P(b, a)^2} \cdot \cdot \cdot \cdot \qquad [2]$$

The right side of Equation [2] is expressed entirely in terms of preference-probabilities, and, although each of these probabilities depends upon the outcomes a and b, Equation [2] states that this particular function of the preference-probabilities is independent of a and b.[3]

The original purpose of the present study was: (a) to make this test in isolation; and (b) to test the whole of the theoretical structure given in the paper cited above. One feature of that structure—as of so many probabilistic models in psychology—is that practically all of the probabilities of a choice must be different from 0 and 1. This leads to doubt about the model at the outset, and it certainly implies that it is wrong when the outcomes are money, because the larger of two sums is always preferred. Nonetheless, we had hoped that we could choose outcomes for which the preferences would not be completely consistent. We selected for this purpose a set of four brands of cigarettes that lay between the fourth and the tenth rank on each *S*'s ordering of 30 popular bands. In addition, we studied gambles with money to see what difference might exist between the two classes of outcome.

---

[2] R. D. Luce, A probabilistic theory of utility, *Econometrica,* 26, 1958, 193-224. See also Appendix 1 of Luce and Raiffa, *op. cit.,* 371-384.

[3] The statistical issue in estimating $Q(\alpha,\beta)$ and in testing whether the right side of Equation 2 is independent of a and b are discussed in Luce, *op. cit.,* 217-219.

As will be seen, the choices of our *S*s exhibit many proportions that are 0 or 1, and so we were not successful in meeting the requirement of the over-all theory. Thus, the complete theory is an inadequate description of the results obtained from these *S*s. A more detailed analysis of the data is needed, however, before we can draw conclusions about the adequacy of the decomposition-axiom.

While the experiment was under way, another, more realistic, theory of choice was developed that rests primarily upon one fundamental axiom.[4] This axiom, when coupled with the decomposition-axiom, leads to certain predictions that can be used to test the theory.[5] Although the present experiment was not designed with this theory in mind, and so we cannot make some of the more powerful tests, it is suitable to check two weaker consequences.

The first is the well-known property of transitivity, namely:

$$\text{if } P(x, y) = 1 \text{ and } P(y, z) = 1, \text{ then } P(x, z) = 1, \cdots \qquad [3]$$

where x, y, and z are gambles. The second is strong stochastic transitivity,[6] *ie.*

$$\text{if } 1/2 < P(x, y) < 1 \text{ and } 1/2 < P(y, z) < 1,$$
$$\text{then } P(x, z) > P(x, y) \text{ and } P(x, z) > P(y, z). \cdots \qquad [4]$$

Strong stochastic transitivity is a generalization of the familiar idea that if x is one *jnd* larger than y and y one *jnd* larger than z, then x will be at least one *jnd* larger than z. In stating this requirement, we do not restrict ourselves to *jnd*-separations, but simply require that if there is a tendency to choose x over y and y over z, then the probability of choosing x over z shall not be less than either of the other two probabilities.

Our primary aim, then, will be to determine for each *S* whether the requirements of the decomposition-axiom, of transitivity, and of strong stochastic transitivity are met. Prior to that, however, it will be necessary to determine whether our *S*s exhibited any significant sequential effects, such as learning, because each choice was presented many times over a period of weeks. The choices of each *S* will, therefore, be analyzed separately.

Method: Subjects. The *S*s (5 in number—1 woman and 4 men) were students of Columbia University; each smoked at least 20 cigarettes per day. They were paid $1.00 for each experimental session and, in addition, they won a package of some brand of cigarettes and a sum of money ranging from 1 cent to 50 cents averaging 24.9 cents per session. The *S*s were naïve regarding the purpose of the experiment.

---

[4] Luce, *Individual Choice Behavior*, 1959, 6.

[5] *Ibid.*, 75-90.

[6] *Ibid.*, 19.

Procedure. On each of the 74 trials of a session E presented S with a choice between two gambles, such as "fifty cents if red occurs and twenty cents if not, or fifty cents if green occurs and one cent if not." A second class of gambles involved cigarette brands as outcomes. After S decided which of the two gambles he preferred, the outcome was determined by E on a pinball machine. The choice and outcome were recorded by E, but S was not paid off until the end of the session. E then selected at random one trial paid in cigarettes and one in money, and S received the recorded outcome.

E randomized the order of presentation of the gambles within each session by selecting, without replacement, numbered chips from a box. These chips were also used to determine which trials were actually paid off. The order in which a gamble was read was reversed from session to session and within any one session half the pairs had the larger expected value read first.

Gambles. Sixty-eight of the 74 choices were of the form: $a\alpha b$ vs. $c\beta d$, where the outcomes, a,b,c, and d, were not necessarily all different. In 34 of these 68 cases, the outcomes were sums of money, the amounts being 1, 10, 20, and 50; in the other 34, they were four brands of cigarette. The events, α and β, were characterized in terms of one of four colors occurring in the pinball machine. The six remaining choices were between pairs of the four cigarette brands.

Sessions. Each session consisted of 74 choices between a pair of gambles. Whenever possible, six sessions were run per week, but there was never more than one per day. The first five sessions were pre-training and they are not included in the analyses. Following that, data were collected for 50 sessions from S_1, S_3, and S_4; for 45 sessions from S_2; and for 32 sessions from S_5.

During the first and at the end of Session 17 and 34, and the final session, S was required to rank order 30 popularly priced types of cigarettes (18 brands, some in variations of regular or king size and of filter or not) and to answer questions about his smoking habits and preferences. On the basis of the first ranking, four brands were chosen for each S from the upper third of his range and these were used as rewards throughout the experiment.

Apparatus. The pinball machine consisted of a 12 × 18-in. tilted board. A glass marble when released at the top of the board would wend its way among the 22 bumpers of the board into one of 10 compartments at the bottom. One roll of the marble would determine the occurrence or non-occurrence of a given event, such as red or not red.

Each compartment was labelled by two different colors. Each of the four colors, red, green, black, and silver, appeared on five compartments. The red event, for example, occurred if the marble dropped into any of the five red compartments; it did not occur if the marble dropped in one of the other five. It will be recalled that a gamble paid off as a consequence of a given event.

The bumpers were so located that the marble entered the compartments with unequal frequencies, thus making it difficult for S to calculate the probability of an event. The colors were assigned in an attempt to produce events with objective probabilities of 0.33, 0.45, 0.55, and 0.67. Prior to the experiment proper, we rolled the marble 500 times to get relative frequencies for the four colors and the stability of the mechanism was checked from 200 randomly chosen trials for each event during the latter half of the experiment. These two sets of relative frequencies are shown in Table I. It will be noted that the largest difference between the first and the second test is about two standard deviations and the other three differences are much less.

Results: (1) Consistency of preferences. The theory proposed states that almost no preferences can be consistent. The percentages of cases for which the observed proportions were 0 or 1 are summarized in Table II. For four of the five *S*s, a substantial fraction, more than 32%, of the choices was perfectly consistent; therefore, we conclude that this theory cannot be applied generally. It was deemed not worthwhile to attempt to check it for the remaining *S*, *i.e.* S_2.

(2) Sequential effects. Any experiment that extends over two and a half months, that involves events with initially unknown probabilities, and that has cigarette brands as outcomes, is an experiment for which it would be unwise to assume without investigation that the choices will have constant probabilities. We shall examine this question in two ways.

TABLE I

RELATIVE FREQUENCIES OF THE FOUR EVENTS ON WHICH THE OUTCOME OF A GAMBLE DEPENDED

Event	500 pre-experimental rolls	200 experimental rolls
α	.672	.615
β	.546	.540
γ	.445	.465
δ	.337	.335

TABLE II

PERCENTAGE OF PAIRS OF GAMBLES FOR WHICH ONE ALTERNATIVE WAS ALWAYS PREFERRED TO THE OTHER

S	Cigarettes (6 cases)	Cigarette gambles (34 cases)	Money gambles (34 cases)
1	83.5	32.4	76.5
2	66.7	0.0	17.7
3	66.7	32.4	55.9
4	100.0	32.4	38.3
5	100.0	67.7	79.5

First, we ask whether there is a consistent trend in the proportion of choices of gambles from the first to the second half of the experiment. Secondly, we ask whether the run-structure of the responses suggests a lack of independence in the choices. In both analyses, we separate the cigarette data into those choices involving at least one pair of brands for which *S*'s preferences were not invariably the same and into those choices that did not involve any inconsistent pair of brands. We expect that experience with the brands may well have modified preferences for a brand of cigarettes over time, and this is most likely to have happened with pairs of brands for which S was initially inconsistent.

Using the data from the first half of the experiment, we estimated the probabilities of each choice. Pairs of gambles with a proportion of 0 or 1 were not considered; all other pairs were taken in the order of the pair which was chosen more than half the time. For these pairs, the proportion of choices was also calculated from the second half of the data. The number of cases for which the proportion from the second half was larger, the same, or smaller was determined and a sign test applied. The resulting p-values are shown in Table III. Where there were inconsistencies in the preference for a brand, there is a significant (5% level) change in the choice of gambles involving these pairs, but no significant effect is found either in the gambles involving consistently chosen brands or in the gambles for money. Thus, apparently, for S_1, S_2, and S_3 there were some changes of preference among the brands, but no systematic changes in their judgments of the relative likelihood of the events used in the gambles. S_4, however, exhibits a significant (1% level) changes in both classes of gambles; the direction is toward greater consistency, suggesting that his estimate of the likelihood of the various colors improved with experience. Finally, S_5 exhibits a significant (5% level) change in the gambles for cigarettes, but not in the gambles for money. One suspects that she altered the relative spacing of her preferences for some brands without, however, affecting her choices of the gambles paid off by cigarettes, which remained consistent.

TABLE III

PROBABILITY, BASED ON A SIGN TEST, THAT THE PROPORTION OF CHOICES CHANGED FROM THE FIRST TO THE SECOND HALF OF THE EXPERIMENT BY CHANCE

The number of cases on which the test is based is shown in parentheses.

	Outcome of gamble		
S	Brands not consistently preferred	Brands consistently preferred	Money
1	.05 (10)	$>.25$ (11)	$>.25$ (8)
2	.05 (18)	$>.25$ (15)	$>.25$ (22)
3	.05 (7)	.25 (11)	.25 (8)
4	—	$<.01$ (22)	.01 (20)
5	—	.05 (10)	.25 (7)

To look into more local effects, we examined the runs of responses to gambles for which the over-all proportion of choices was different from 0 and 1. The results are summarized in Table IV. S_2, S_3, and S_5 gave runs of responses above chance (5% level) to slightly more pairs than one would expect. For example, S_2's gambles with payoffs in money led to 3 significant cases out of 28, whereas by chance there should have been 1.4. The discrepancies are so small, however, that we doubt that there were serious deviations for these *S*s. S_4, whose choices differed in the first and second halves of the experiment, also exhibits a significant number of runs in more than the expected number of cases. So does S_1. In almost all cases, there were too few rather than too many runs, thus suggesting that there was a perseveration effect.

Although there are some deviations in the responses of our *S*s from a model of independent Bernoulli observations, there are still enough cases where this model appears adequate to warrant analysis of the data in terms of the three hypotheses that were discussed earlier.

(3) Transitivity. For both the brands and money, there were 23 distinct triplets of gambles in which *S* made choices between all three pairs. Of the data for these triplets, some satisfy the hypotheses of Equation [3] (transitivity-property) provided that estimates $\hat{P}$ are substituted for probabilities *P*, some appear to satisfy the hypotheses of Equation [4] (strong stochastic transitivity) and the rest satisfy neither set of hypotheses and will be ignored. The data satisfying the hypotheses of the transitivity-property can be summarized as shown in Table V. Of the 56 cases involving money, only one violates transitivity. Of course, this may not be an actual violation because the observed 0 and 1 proportions may not, in fact, signify 0 and 1 probabilities. We conclude that transitivity cannot be rejected for money-gambles.

TABLE IV

THE NUMBER OF PAIRS OF GAMBLES THAT EXHIBITED A SIGNIFICANT (5% LEVEL) NUMBER OF RUNS

	Cigarettes		Money	
S	No. with more than one run	No. with a significant deviation	No. with more than one run	No. with a significant deviation
1	8	6	23	5
2	36	1	28	3
3	25	3	15	1
4	23	3	21	5
5	11	1	7	1

The picture is less favorable for transitivity in the cigarette data, particularly for S_1 and S_3. It is interesting, however, that all the violations for these two *S*s involve at least one pair of brands for which *S* did not have a consistent preference. As we noted earlier, there were changes over time in the data for these gambles, and so the test of transitivity may not be legitimate.

(4) Stochastic transitivity. Parallel data for a test of strong stochastic transitivity are given in Table VI. The hypotheses underlying such a test were only met in the data for gambles paid off with cigarettes. There are 12 apparent violations in 47 cases. It is not easy to know how to test this assumption because, even if it were true, we would expect some apparent violations to arise from sampling errors. No one has yet worked out the sampling distribution, even for particular assumptions about the dependence of $P(x,z)$ upon $P(x,y)$ and $P(y,z)$.

TABLE V

NUMBER OF VIOLATIONS OF THE ASSUMPTION OF TRANSITIVITY

	Cigarettes		Money	
S	No. cases	No. violations	No. cases	No. violations
1	2	2	13	0
2	0	0	13	1
3	3	3	5	0
4	2	0	4	0
5	13	1	21	0

TABLE VI

NUMBER OF VIOLATIONS OF THE STRONG STOCHASTIC ASSUMPTION OF TRANSITIVITY IN GAMBLING FOR CIGARETTES

S	No. cases	No. violations	*P* under the null hypothesis
1	6	2	.100
2	23	8	.002
3	7	1	.007
4	10	1	.000
5	1	0	.333

We may show, first, that the data exhibit some structure. As a null hypothesis, suppose that each of the six orderings of the three proportions of choices is equally likely, then two of the six will be consistent with strong stochastic transitivity. The binomial probabilities of our observations under this hypothesis are shown in Table VI, and it is clear that we must reject this null hypothesis.

Looking at the violations more carefully, two points seem important. First, of the 12 violations, only 3 (one for S_1 and two for S_2) are violations to the extent of more than two standard deviations of the estimated probabilities. Second, of the 11 violations exhibited by S_1, S_2, and S_3, only one does not involve a pair of brands for which that *S* was inconsistent. As we have seen earlier, the stability of the probabilities of choice between such gambles is in doubt.

On the basis of these considerations, we conclude that these data are not adequate to reject the strong stochastic transitivity-condition; but we would not claim that the support for it is strong.

(5) Decomposition-axiom. The decomposition-axiom, in the form of Equation [2], is used to estimate the postulated judgment of likelihood of events, $Q(\alpha,\beta)$, from the proportions of choices of gambles, $\hat{P}$. For each outcome pair (a,b) for which we have estimates of the *P*-terms, Equation [2] gives us an estimate of $Q(\alpha,\beta)$. The means of these estimates are shown in Table VII. If the axiom is correct, the several estimates of each *Q* should be approximately the same. To test this, we take the over-all mean as the expected value, convert all of the proportions to frequencies, and do a χ^2 test. Only gambles involving money and cigarette brands that were consistently preferred have entered into these calculations.

TABLE VII

VALUES OF *Q* ESTIMATED FROM THE DECOMPOSITION-AXIOM (EQUATION [2])
Events are denoted by Greek letters.

S		Cigarettes			Money		
		β	γ	δ	β	γ	δ
1	α	.970	1.000	.960*	.996	1.000	1.000
	β		.760	.990		.770	.980
	γ			.955			.960
2	α	.363	.567*	.933	.213	.589	.956
	β		.667	.889		.856	.889
	γ			.889†			.960
3	α	.990	.880	.990	.984	.980	.996
	β		.880	.980		1.000	.990
	γ			.965			.988
4	α	.912	.980	.976	.928	.980	.968
	β		.870	.910		.810*	.940
	γ			.832			.852*
5	α	.956†	1.000	.938†	.994	1.000	1.000
	β		.969	.891†		.985	1.000
	γ			.856			.769

* Significant at 5% level. † Significant at 1% level.

Of the ten subtables of Table VII, seven do not have an over-all χ^2 significant at the 5% level. The S_1-cigarette table is significant at the 5% level. The S_4-money and the S_5-cigarette tables are significant at the 1% level. It will be recalled that we found some indications of instability and dependence of responses in the choices of S_4 and of S_5 when the gambles are paid off by cigarettes (Table III), in which cases the test of the decomposition-axiom is not suitable. Ignoring the S_4- and S_5-cigarette subtables (with their 1% levels), we note that the remaining eight tables include 48 estimates. Of these, two are significant at the 5% level and one at the 1% level, which is about what one would expect by chance. This evidence appears to support the decomposition-axiom, at least for small payoffs when the preferences are stable over time.

One point of interest is S_2's reversal of the relative likelihood of the events (colors) α and β from the order given by the other *S*s and from their objective probabilities of occurring. It is a little difficult to see how he could have misjudged which event was more likely after hundreds of observations, but he did.

SUMMARY

A gambling experiment employing both money and cigarettes as the outcome was run to test three assumptions now current in a stochastic theory of choice. Five *S*s chose between 74 pairs of gambles in each of from 32 to 50 daily sessions. Probabilities of choice were estimated from these data for each *S*.

Some sequential changes in the choices of the *S*s were established. For one *S*, they appeared about equally in both the cigarette and money data, suggesting that he altered his judgment of the relative likelihood of the events during the experiment. For the other four *S*s, sequential effects appeared mainly in the cigarette data, and they seemed to be closely correlated with changes of brand preferences. It was found that gambles involving inconsistently preferred brands accounted for most of the sequential effects. There was no evidence that these *S*s altered their judgments about the likelihood of the events on which the outcome of the gamble depended.

Ignoring those pairs of gambles involving at least one pair of inconsistently preferred brands, we found strong support for the transitivity-assumption and more modest support for the strong stochastic transitivity- and decomposition-assumptions.

THE EFFECT OF HUE AND BRIGHTNESS ON THE SIZE-ILLUSION OF CONCENTRIC CIRCLES

By TADASU OYAMA, Hokkaido University, Japan

In only a few of the many studies made on geometrical, optical illusions has the effect of hue been investigated. Benussi was the first to do so and he found in his study of the Zöllner illusion that hue had little effect on the illusion and that brightness had a differential effect when the figure was drawn in two colors which differed in brightness. When the vertical lines were of one color and the transverse lines of another, the illusion *increased* as the brightness-difference between the transverse lines and the background increased, and *decreased* as the brightness-difference between the vertical lines and the background increased.[1]

If these results are corroborated by other experimenters and with other illusions, they will give us an explanation of the illusions of this type. In particular, they will decide the issue between the physiological theory of neural induction and the Gestalt theory that similarity is a decisive factor in phonemonal grouping.

The size-illusion of concentric circles (see Fig. 1), described by Delboeuf in 1893,[2] was selected for investigation because it has been used in recent years in numerous studies—particularly in Japan.[3] The inner circle of B in Fig. 1 appears larger than Circle A which is physically the same in size; and the outer circle of B appears small than Circle C which again is physically the same in size.

In working with this illusion, Morinaga found that over-estimation was maximal when the ratio of the diameters of the concentric circles was 3:2.[4] This ratio was confirmed by Ogasawara under three conditions of absolute size and by Morinaga and Kansaku under various conditions of brightness.[5] Regarding the effect of brightness on this illusion, Wada obtained results similar to those of Benussi when achromatic circles of 3:2 ratio were used and their brightnesses were varied independently on white and black backgrounds;[6] and Morinaga and Kansaku, with concentric circles of various size-ratios, obtained similar results with a gray background.[7]

---

* Received for publication September 15, 1960. The writer is indebted to Professors S. Moringa, Y. Wada, and Y. Takada for their valuable suggestions and to Mr. S. Fujii and Miss R. Akatsuka for their assistance in collecting the data.

[1] Vittoris Benussi, Ueber den Einfluss der Fabre and Grösse der Zöllnerschen Tãuschung, *Z. Psychol.*, 29, 1902, 264-351, 385-433.

[2] J. L. R. Delboeuf, Sur une nouvelle illusion d'optique, *Rev. scient.*, 51, 1893, 237-241.

[3] Tadasu Oyama, Japanese studies on the so-called geometrical-optical illusions, *Psychologia*, 3, 1950, 7-20; S. Morinaga and H. Kansaku, A study on the influence of the figural lightness upon the illusion of concentric circles. *Jap. J. Psychol.*, 32, 1961, 148-159. Y. Wada, Studies on geometrical illusions: III. The effect of intensity-difference, *Proc. 20th Annual Meeting Jap. Psychol. Assn.*, 1956, 62-63; J. Ogasawara, Displacement-effect of concentric circles, *Jap. J. Psychol.*, 22, 1952, 224-234; S. Morinaga, The conditions of the size-assimilation and size-contrast, in K. Hirose (ed.) *Masuda Hakushi Shaon Saikin Shinrigaku Ronbunshu*, 1935, 28-48.

[4] Morinaga, *op. cit.*, 28-48.

The present study was undertaken to analyze further the effect of hue and brightness upon the size-illusion of concentric circles. Of especial interest was the discovery of the effect of similarity and dissimilarity when the circles were drawn in the same or different hues and brightnesses. For example, if the solid lines in Fig. 1 are of one color and brightness and the broken lines are of another color and brightness, then, if similarity is a favorable condition, the illusion should be greater in B than in D and in E than in F.

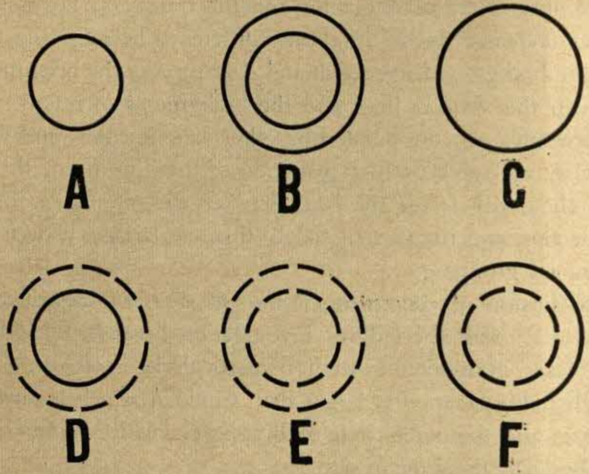


FIG. 1. STIMULUS-FIGURES FOR ILLUSION OF CONCENTRIC CIRCLES.
Solid lines represent one color and broke lines another color.

Method: Apparatus. A wooden booth, 80 cm. high, 90 cm. wide, and 115 cm. deep, was constructed. The inner surfaces—the sides, top, and bottom—were painted a medium gray whose brightness was approximately 6.5 on the Munsell scale. The back end was covered, in Experiments I and II, with white cardboard of 9.0 brightness and, in Experiments III and IV, with gray cardboard of 7.5 brightness.

S sat in a chair before the booth and observed the stimulus-figures binocularly as they were exposed in two apertures cut into the end screens. These apertures were square (20 × 20 cm.) and they were centered in the screen at the level of *S*'s eyes. The right lower corner of the upper aperture was 10 cm. above and 10 cm. to the left of the upper left corner of the lower aperture.

---

[5] Ogasawara, *op. cit.,* 224-234; Morinaga and Kansaku, *op. cit.,* 148-159.
[6] Wada, *op. cit.,* 62-63.
[7] Morinaga and Kansaku, *op. cit.,* 148-159.

The standard stimulus-figure was exposed in the upper aperture and the comparison-figure in the lower aperture. In Experiment I, the screen was illuminated by three 60-v. daylight bulbs. Two of these lights were placed above and slightly behind *S* and the third in front, below his chin. In Experiments II-IV, these bulbs were replaced by a Standard Illuminator,[8] matched with C.I.E. Standard Source C (6500° K), which was placed close to the right of *S*. The illuminations under these conditions were 5 ft.c. and 2 ft.c., respectively.

Stimulus-materials. Three types of stimulus-materials were used: standard, control, and comparison. The standard consisted of two concentric circles cut from standard papers which were pasted on white cards. The circular outlines were 2 mm. wide and the diameters of the circles, as measured from the middle of their outlines, were 90 and 60 mm. (ratio of 3:2), respectively. At *S*'s observational distance, the width of the outlines subtended a visual angle of 6′ and the diameters of the circles, visual angles of 4.5° and 3°, respectively.

The control-card, which was similarly prepared, contained a single circle, the inner of the two on the standard cards. The comparison-stimuli consisted of a series of 25 white cards on which black outline circles, drawn in India ink, were centered. The diameters of these circles varied, in steps of 1 mm., from 48 to 72 mm.

Procedure. From four to eight experimental sessions were held for every *S* in each of the experiments reported in this study. *S* was instructed in the different experiments (1) to compare the apparent size of the inner circles of the standard stimulus-cards with the circles of the comparison-cards; and (2) to compare the circle on the control-card with those on the comparison-cards. The method of limits, with two asecending and two descending series, was used to determine the point of subjective equality (*PSE*). *S*'s reports were limited to 'larger,' 'smaller,' or 'doubtful.' The difference between the *PSE*s under these two conditions was taken as the measure of the illusion.

Subjects. Staff members and students in the Department of Psychology in numbers later reported served as the *S*s during the several experiments. Although some of the *S*s wore glasses, all had normal vision—color and acuity.

EXPERIMENT I

Experiment I was designed to determine the effects of brightness upon the illusion. Though this problem, as noted above, has previously been studied, we chose to reinvestigate it because in none of the earlier studies was brightness systematically varied. In the present experiment, all of the stimulus-cards were achromatic.

Procedure. The standard stimulus-materials were of two kinds: a black series (Bk) and a gray series (Gr). Each series contained 11 stimulus-cards and one control card. In the Bk-Series, the inner circles were black (1.5 Munsell value) and the outer circles were of 11 different brightnesses as follows: 1.0, 2.0, 2.5, 3.0, 4.0,

---

[8] This illuminator was obtained from the Japanese Color Research Institute.

4.7, 5.3, 6.0, 7.0, 8.0, and 9.3.[9] In the Gr-Series, the inner circles were gray (4.7 Munsell value) and the outer circles were of the same brightnesses as in the Bk-Series. The control-card for each series was the same in size and brightness as the inner circle of the standard series, *i.e.* black (1.5 Munsell value) in the Bk-Series, and medium gray (4.7 value) in the Gr-Series. The comparison stimulus-cards were as described above.

Four *S*s served in eight sessions; four sessions with each series. In every session, the comparison-cards were exposed in ascending or descending order with every standard card and the control-card of each series. This procedure was repated in four sessions, hence two ascending and two descending series of comparisons were obtained from which the *PSE*s of every standard stimulus-card and every *S* were computed.

Results. The solid line in Fig. 2A shows the average magnitude of the illusion in the Bk-Series and the broken line that in the Gr-Series. The two curves are strikingly similar. They are flat when the outer circle varies in brightness from 1.5 to 6.0 and begin to show a decrease when the brightness reaches 7.0 from whence they decrease rapidly to 9.3.

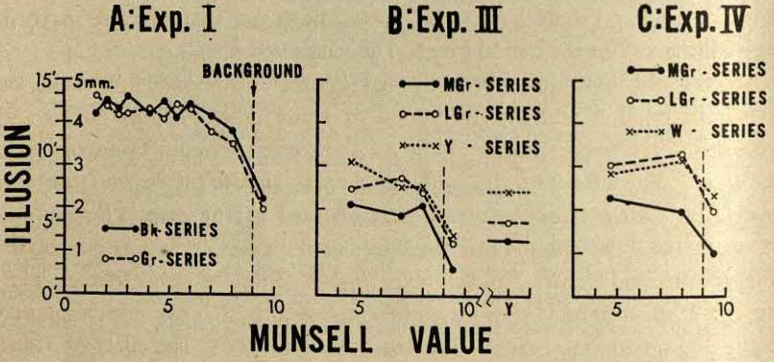


FIG. 2. MAGNITUDE OF THE ILLUSION AS A FUNCTION OF THE BRIGHTNESS OF THE OUTER CIRCLE

These results mean that the brightness of inner circles used in this experiment had little effect on the illusion, and that the illusion remained at a constant level when the brightness-difference between the outer circle and the background was greater than a certain value. It should be noted that the brightest outer circle in this experiment was slightly brighter than the background and that the absolute magnitude of brightness-difference seemed to act in the same manner even when the direction of difference was changed from negative to positive.

---

[9] These values in Munsell notation were determined by comparison with the Japanese *Standard Chart of Colors.*

The analysis of variance of these results indicates that the effects of the outer circle (*OC*), the subject (*S*), and the interaction between the inner circle (*IC*) and the subject ($IC \times S$) were statistically significant when the secondary interaction ($OC \times IC \times S$) was adopted as the error, while the effects of the inner circle (*IC*) and the other primary interactions ($IC \times OC$ and $OC \times S$) were insignificant. The results of this analysis are in line with these shown in Fig 2A. The insignificance of the effect of the inner circle does not necessarily mean any general tendency, because only two levels of brightness of the inner circle were used and there remains a possibility that a smaller brightness-difference from background will produce different results. If the similarity in brightness between the outer and the inner circles was a determining factor of this illusion, the illusions of the black and gray inner circles should be affected in different manners by the brightness of the outer circles. The interaction between the inner and outer circles ($IC \times OC$) was, however, insignificant. Accordingly, the hypothesis of similarity was not supported.

Because the physical errors in making the stimulus-figures have no systematic relation with the brightness of either outer or inner circles and were absolutely independent of the *S*s, their effects should be involved in $IC \times OC$. The effect of the outer circle was significant beyond the 1% level again when this interaction was adopted as the error ($F = 24.06$, $df = 10/10$). This means that the effect of the outer circle was not ascribable to physical errors. In this experiment, different inner circles were presented in different sessions, and consequently $IC \times S$ was supposed to involve session-to-session variations in individual *S*s. The effect of *S* was also significant beyond the 1% level when this interaction was adopted as error ($F = 100.81$, $df = 3/3$).

EXPERIMENT II

In Experiment II, four chromatic colors, red, yellow, green, and blue, and gray were used in both the outer and inner circles, and the effect of hue on the illusion was examined.

Procedure. Five series of stimulus-figures, R, Y, G, B, Gr, were used. Each series consisted of five illusional figures and a control-figure. The inner circles were red (5.0 R, 4.0/12.0 in Munsell notation) in the R-Series; yellow (7.5 Y, 8.3/11.0) in the Y-Series; green (5.0 G, 5.5/8.0), in the G-Series; blue (2.5 PB, 4.0/10.0) in the B-Series; and medium gray (4.7) in the Gr-Series. The outer circles in each series were made from the same five colors.

Four new *S*s served in eight sessions, four with the G- and B-Series and four with the R-, Y-, and Gr-Series. In every session, all figures of the two or three series were used, and one ascending or descending series of comparison-stimuli were used for each figure.

Results. Average illusions in the 25 figures used in this experiment are shown in Table I, in which each row represents each series (inner circle) and each column represents each outer circle. An inspection of this table reveals that the illusions in the Y- and G-Series were slightly larger than in the other series and that the illusions of R- and B-Series were nearly equal to those of Gr-Series, and that the illusions caused by the yellow outer circle were smallest in every series and the illusions caused by the outer circles of the other hues were approximately on the same level as those of medium gray. There seems to be no peculiarity in the illusions of the figures which consisted of the outer and inner circles of an identical color.

TABLE I

ILLUSIONS AS A FUNCTION OF COLORS OF INNER AND OUTER CIRCLES

(Experiment II)

Inner circle	Outer circle					
	Red	Yellow	Green	Blue	Gray	Av.
Red	4.3 mm	3.5	4.0	4.4	3.7	4.0
Yellow	4.4	4.2	4.6	4.6	4.4	4.4
Green	4.4	3.2	4.9	4.6	4.4	4.3
Blue	4.1	2.5	4.2	4.2	4.3	3.9
Gray	4.0	3.2	4.0	3.9	3.8	3.8
Av.	4.2	3.3	4.3	4.3	4.1	4.05

The analysis of variance indicates that the effects of IC, OC, S, and $IC \times S$ were significant, while $IC \times OC$ and $OC \times S$ were insignificant, when $IC \times OC \times S$ was used as the error. IC and OC were significant beyond the 1% level and so also when $IC \times OC$ was used as the error (Fs of 5.45 and 11.81, with dfs of 4/16 and 4/16). S was also significant beyond the 1% level when $IC \times S$ was adopted as the error ($F = 27.05$, $df = 3/12$), while IC was insignificant in that case. The difference between the yellow outer circles and either green or blue outer circles was significant beyond the 5% level by Tukey's test using the Studentized range table when $IC \times OC$ was used as the error.[10] No other difference among outer circles or inner circles was significant in this test.

These analyses suggest that the illusion was changed systematically by the variations of the colors of the outer and inner circles, but it was not affected by the similarity or dissimilarity between these colors. The effects of red, blue, and green were nearly equal to the effects of the medium gray with similar lightness, in both cases in which these colors were used in the outer and inner circles, except that green inner circles resulted in slightly (insignificantly) larger illusions. The statistical significance in the differences of illusion due to the colors is supposed to be caused mostly by yellow, because it brought about the different results from the other colors when it was used as the color of either outer or inner circles. Yellow had isolated value in lightness from the others, and its peculiarity in results may be attribued to its specialty in lightness. This hypothesis will be tested in the next experiment.

---

[10] J. W. Tukey, Quick and dirty methods in statistics, II. Simple analysis for standard designs, *Proc. 5th Annual Convention Amer. Soc. Quality Control*, 1951, 189-197. Cited in H. Scheffé, A method for judging all contrasts in the analysis of variance, Biometrika, 40, 1953, 87-104.

EXPERIMENT III

Experiment III was designed to analyze the peculiarity of yellow observed in Experimental II. The possibility, suggested in Experiment I, that smaller brightness-difference of the inner circle from background will produce greater illusions will be also examined in this experiment.

Procedures. Three series of stimuli, each of which consisted of four illusional figures and a control-figure, were used. The inner circles were medium gray (4.7) in MGr-Series; light gray (8.0) in LGr-Series, and yellow (7.5 Y. 8.3/11.0) in Y-Series. The outer circles were gray with various lightnesses (4.7, 7.0 and 8.0) and yellow.

Three new *S*s and a *S* who had served in Experiment II were used in six sessions; two for each series. In each session, all figures of a series were used, for each figure one ascending and one descending. In regard to the last point, this experiment was different from Experiments I and II.

Results. Three curves in Fig. 2 B represent the average illusions caused by various gray outer circles in three series respectively. Three short horizontal lines in the right end of the figure show the illusions caused by yellow outer circles in the three series. Three curves are similar in shape to each other and to the curves in Fig. 2 A, but lower in general level than the curves in Fig. 2 A. The solid line in Fig. 2 B is distinctly lower than the broken line in Fig. 2 A despite the fact that they represent the results obtained under practically the same conditions. This difference is attributed to the differences among the *S*s, since individual differences in such illusions are large, as shown by Cymbalistyj.[11] In Fig. 2 B, the Y-Curve is highest, the LGr-Curve the next, and the

---

[11] B. J. Cymbalistyj, Grössenangleichung und Grössenkontrast in den geometrisch-optischen Täuschungen, *Psychol. Forsch.*, 23, 1949, 115-184.

MGr-Curve lowest. Each horizontal line is on a level with a point between 8.0, and 9.3 of the corresponding curve. This suggests that the effect of the yellow outer circle is smaller than the effects of the gray outer circles whose lightnesses are 8.0, and larger than those of gray whose lightness is 9.3.

The analysis of variance made here shows that, when $IC \times OC \times S$ was used as the error, IC, OC, S, $IC \times S$ and $OC \times S$ were significant, but $IC \times OC$ was insignificant again. IC and OC were significant beyond the 1% level also when $IC \times OC$, which would involve physical errors, was used as the error (Fs of 17.00 and 19.95, dfs of 2/8 and 4/8). OC and S were significant beyond the 1% level, when $OC \times S$ was used as the error (Fs of 6.32 and 25.42, dfs of 4/12 and 3/12). When $IC \times S$ was used as the error, S was significant beyond the 5% level ($F = 9.19$, $df = 3/6$), but IC did not reach the significant level. Tukey's test indicated that both the differences between Y- and MGr-Series and between the lightest outer circle and the others were significant beyond the 5% level when $IC \times OC$ was used as error, but no other difference among the outer or inner circles reached the significant level.

As shown by these analyses, yellow and light gray, when used as the color of either outer or inner circle, did not yield significantly different results. In this respect, the hypothesis mentioned above, that the peculiarity of yellow found in Experiment II is ascribable to its lightness, was supported. The second hypothesis tested in this experiment, that greater illusions will be found in lighter inner circles, was not supported because the difference between the MGr- and the LGr-Series was not statistically significant by Tukey's test. The difference, however, was only slightly less than the critical value, and it is quite probable that a greater number of Ss or greater difference in the brightness would result in statistical significance.

EXPERIMENT IV

In Experiment IV, the effect of brightness of the inner circle was examined again. White inner circles were added to those previously used,

Procedures. Twelve figures divided into three series were used. They were all achromatic. The inner circles were 4.7 in the MGr-Series; 8.0 in the LGr-Series; and 9.3 in the W-Series. The outer circles were the same three achromatic colors. One control-stimulus without outer circle was included in each series.

The Ss used in Experiment III served in four sessions. In each session, all the figures, with one ascending or descending series of comparison-circles were used.

Results. The average illusions, shown by three curves in Fig. 2 C, indicate effects of brightness of the outer circles similar to those obtained in Experiments I and II. The level of LGr-Series was somewhat higher than that of the LGR-Series of Experiment III, despite the fact that the same figures and the same *S*s were used, while the results of the MGr-Series were practically the same as those obtained in Experiment III.

Analysis of variance showed that the effects of *IC*, *OC*, and *S* were significant when $IC \times OC \times S$ was used as the error. The effects of *IC* and *OC* were significant beyond the 5% level when $IC \times OC$ was used as the error (*F*s of 15.96 and 15.48, with *df*s of 2/4 and 2/4).

These results provide positive evidence for the hypotheses previously mentioned: (1) that the size-illusion increases as the brightness-difference of the inner circle from the background is reduced; and (2) that the peculiarity of the yellow inner circle should be attributed to its small brightness-difference from the background.

DISSCUSSION

(1) One of the most important findings of this study is that the interactions between the outer and inner circles were statistically insignificant in all the experiments. This means that there was no effect of similarity or dissimilarity in hue and brightness between the two circles. Inspection of Fig. 2 and of Table 1 corroborates this conclusion. In the comparison of concentric circles of different colors, no effect on the size-illusion was discovered. These results have the following theoretical implications.

(a) If this illusion is considered as an example of physiological induction in retinal or cortical fields, as hypothesized by Motokawa and by Köhler and Wallach, the induction should be regarded as independent of the similarity between the physiological correlates of the colors of the inducing and the induced parts.[12]

(b) If the Gestalt principles of Wertheimer are applied to this illusion, the principle of similarity should play an important role and the illusion should be greater when the concentric circles are of the same color.[13] This expectation, however, was not sustained. A simple application of Gestalt principles in the explanation of this illusion is, therefore useless.

(c) If the phenomenal character that the concentric circles form a ring or dough-nut shaped object is essential to the illusion, as Cymbalistyj and Ogasawara hold,[14] then the similarity between the circles should be a determining factor in the illusion. The findings of this study are contrary to this expectation and disprove the explanation.

---

[12] K. Motokawa, Field of retinal indiction and optical illusion, *J. Neurophysiol.*, 13, 1959, 413-426; Wolfgang Köhler and Hans Wallach, Figural after-effects: An investigation of visual processes, *Proc. Amer. phil. Soc.*, 88, 1944, 269-357.

[13] Max Wertheimer, Untersuchungen zur Lehre von der Gestalt: II. *Psychol. Forsch.*, 4, 1923, 301-350.

[14] Cymbalistyj, *op. cit.*, 115-184; Ogasawara, *op. cit.*, 224-234.

(2) The second important finding of this study concerns the effect of brightness-difference from background. The effect of the brightness-difference of the outer circle was statistically significant in Experiments I, III, and IV, and that of the inner circle was significant in Experiment IV. As the brightness-difference of the *outer* (inducing) circle from the background increased, the illusion *increased* rapidly to an asymptote, but further increases in the brightness-differences brought about little or no changes in the illusion. As the brightness-difference between the *inner* (test) circle and the background increased, the illusion *decreased* to a specific level and remained nearly constant at that level. These results support Benussi's finding with Zöllner's figure and Wada's finding with concentric circles. The weak points of the present experiments were (1) that only a white background was used, which restricted the brightness-difference in a negative direction from the background, and (2) that both inner and outer circles, which were found significantly different in their effects from the others, were white and were special cases—not only in the smallness of the absolute magnitude of the brightness-difference from the background but also in the positivity of the direction of difference.

These points are, however, reinforced by the results of Wada's experiments in which both white and black backgrounds were used, and in which the brightness-differences were always in one direction on each background. Wada used as stimulus-figures every combination of a black (1.7) or light gray (8.3) inner circle and a black or light gray outer circle on white background (9.2), and found statistically significant differences between the overestimation of the two kinds of inner circle and between the effects of the two kinds of outer circle. He also found similar results using white (9.1) and dark gray (3.7) circles on black background (2.5).

These results and ours may be summarized, using the concepts of Orbison's field-theory of vectors, as follows:[15] The cohesive forces, which are postulated to act in the inducing parts to attract the test-parts of the figure, increase rapidly as the brightness-difference of inducing parts from background increases up to a certain low level and then keeps a constant value. The restraining forces, which are hypothesized not to act on test-parts to displace the loci of stimulation, increase just in the same way as the brightness-difference of test-parts from background increases.

(3) The third important finding was the ineffectiveness of hue. In Experiments II and III, in which chromatic figures were used, red, yellow, green, and blue outer and inner circles did not result in statistically significant difference from gray circles which matched with them in bright-

---

[15] W. O. Orbison, Shape as a function of the vector-field, this JOURNAL, 52, 1939, 31-45.

ness. To this extent, the effects of colors in the illusion can be attributed solely to the effects of brightness-difference from the background. It may, therefore, be concluded that hue has no effect on this illusion. Koffka and Harrower's classification of hard and soft colors is not applicable in this situation.[16]

(4) The last important finding concerns individual differences. Differences among the *S*s were highly significant in all experiments. This finding confirms Cymbalistyj's results with this and other illusions. Individual differences, however, are concerned with the absolute level of the illusion, rather than with the effectiveness of the outer or inner circles. The statistical significance of the interaction between the inner circle and *S* may chiefly be attributed to session-to-session variations. The interaction between the outer circle and *S* was slightly significant only in Experiment III.

SUMMARY

The overestimation of size of a concentric circle was measured by the method of limits, under various conditions of hue and brightness. The main results obtained were as follows.

(1) The similarity or dissimilarity in hue and brightness between two concentric circles has no effect on this illusion. Consequently neither the Gestalt principle of similarity nor any explanation based on the phenomenological characteristic of "ring" or "doughnut" was useful in the explanation of this illusion.

(2) The illusion *increases* as the brightness-difference of the *outer* (inducing) circle from background increases, and it *decreases* as the brightness-difference of the *inner* (test) circle from background increases. Further increases in the brightness-differences beyond a comparatively small zone did not affect the illusion.

(3) Hue does not have an effect *per se.* Whatever effect it may have is attributable solely to the effect of brightness-difference from background.

---

[16] Kurt Koffka and M. R. Harrower, Colour and organization, *Psychol. Forsch.*, 15, 1931, 145-192; 193-275.

MONOCULAR AND BINOCULAR REPRODUCTION OF BINARY STIMULI APPEARING RIGHT AND LEFT OF FIXATION

By E. RAE HARCUM and DOROTHY W. DYER,
College of William and Mary

In this study, we seek to determine the observer's accuracy in reproducing stimulus-patterns presented in the right and left halves of the visual field. Of particular interest are eye-dominance and hemispherical dominance of the cerebrum, which might account for individual differences in this task.

Differences between the right and left fields of vision have frequently been demonstrated. For example, Dallenbach and his co-workers reported differences in the relative attensity or vividness of test-patches presented to the left or right of fixation.[1] These differences were apparently related to handedness, *i.e.* dominant cerebral hemisphere, with normally the more attense patches left of fixation for right-handed *O*s. In other words, stimuli presented to the non-dominant hemisphere were found to be more attense. With monocular vision, differences in sensitivity, if found, usually favor objects in the temporal fields.[2] Experiments on perceptual span generally reveal superior performance for those stimulus-elements to the left of fixation when elements have appeared simultaneously on both sides of fixation.[3] This difference between right and left visual fields is greater for *O*s with dominant right

---

* Received for publication August 25, 1960. This investigation was supported by research grants (B-1971 and B-2661) from the Institute of Neurological Diseases and Blindness, Public Health Service.

[1] K. M. Dallenbach, Position vs. intensity as a determinant of clearness, this JOURNAL, 34, 1923, 282-286; R. S. Burke, and K. M. Dallenbach, Position vs. intensity as a determinant of attention of left-handed observers, *ibid.*, 35, 1924, 267-269; A. M. White, and K. M. Dallenbach, Position vs. intensity as a determinant of attention of left-handed observers, *ibid.*, 44, 1932, 175-179.

[2] L. L. Sloan, Rate of dark adaptation and regional threshold gradient of the dark-adapted eye: Physiologic and clinical studies, *Amer. J. Ophthal.*, 30, 1947, 705-720.

[3] H. R. Crosland, Letter-position effects, in the range of attention experiment, as affected by the number of letters in each exposure, *J. exp. Psychol.*, 14, 1931, 477-507; I. H. Anderson, The effect of letter-position on range of apprehension scores, with special reference to reading disability, *Univ. of Mich. Sch. of Educ. Bull.*, 18, 1946, 37-40; Woodburn Heron, Perception as a function of retinal locus and attention, this JOURNAL, 70, 1957, 38-48; E. R. Harcum, Three inferred factors in the visual recognition of binary targets, in J. W. Wulfeck and J. H. Taylor (ed.), Form discrimination as related to military problems, *NAS-NRC Publ. 561,* 1957, 32-37; Visual recognition along four meridians of the visual field: Preliminary experiments, *Univ. Mich., Engng. Res. Inst., Proj. Michigan Rept., 2144-50-T,* 1957, 1-15; II. Nine-element typewritten targets, *ibid., 2144-293-T,* 1958, 1-21; E. R. Harcum, and Ausma Rabe, III. Patterns of blackened circles in an eight-circle template, *ibid., 2144-294-T,* 1958, 1-68.

eyes than for those with dominant left eyes; and it is greater for superior readers than for inferior readers.[4]

The problem of the present study is: (1) to determine whether the stimulus-objects to the left or right of fixation are more accurately reproduced, *i.e.* which position is more effective in attracting attention; and (2), since discrepant results have been obtained, to discover the cause of the differences.[5]

The present investigation follows specifically from the theoretical positions of Hebb, Heron, and Harcum.[6] The ability to recognize visual stimuli reflects, as we assume, the functional superiority of their perceptual organizations. If the visual objects are too numerous or too complex to be perceived in their entirety, the component elements must be attended to in some sequence, starting at some reference point. This may require a scanning of the visual field even when the eyes are fixed. The sequence is thought to proceed in the manner of eye-fixations if eye-movements could be made. It is further believed that the areas in the visual field covered first by this scanning will produce fewer errors in report—a *priamcy-effect.* This effect is a general behavioral attribute and not specifically a mechanism of visual perception. The selection of a beginning reference-point in the right or left visual field for the scan may be determined by the perceived clarity of the stimuli. Dallenbach suggested that the distribution of attention may produce such greater clarity for the stimuli in the left visual field for most *O*s.[7]

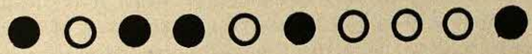


FIG. 1. VISUAL PATTERN USED IN THIS STUDY

Specifically, this study tests the following hypotheses: I. Fewer errors in reproduction occur for stimuli having greater contrast from the background. II. The viewing eye *per se* is not an important factor in relative accuracy for stimulus-elements left or right of fixation, and neither is its interaction with differential stimulus-contrast right or left of fixation. III. Most *O*s exhibit a progressive increase in errors beginning consistently at one end of the target and proceeding toward the opposite end of the target (because of the primacy-effect). IV. For most of our *O*s, this effect favors the elements to the left of fixation. V. There is a tendency for primacy to favor the end of the target containing the stimuli having the greater contrast.

---

[4] Crosland, Superior elementary-school readers contrasted with inferior readers in letter-position, "range of attention," scores, *J. educ. Res.,* 32, 1939, 410-427.

[5] Harcum, Visual recognition along various meridians of the visual field: IX. Monocular and binocular recognition of patterns of squares and circles, *Univ. Mich., Engng. Res. Inst., Proj. Michigan Rept. 2144-307-T,* 1958, 1-19.

[6] D. O. Hebb, *The Organization of Behavior,* 1949, 49-50, 60-106; Heron, *op. cit.,* 45-48; Harcum, *op. cit., NAS-NRC Publ. 561,* 1957, 35-37.

[7] Dallenbach, *op. cit.,* 285-286.

Method. A Dodge tachistoscope was set to expose the stimulus for about 100 m.sec. at a distance of 26 in. The illumination on the target-field and on the fixation-field was produced by ordinary incandescent lamps. Field-luminance under continuous illumination was about 2 ft.-L.

Individual test-patterns were formed by variously blackening or darkening some of the circles in a 10-circle horizontal linear pattern, as in Fig. 1. Twenty patterns of 5 filled circles consisted of 10 arbitrary patterns and their mirror images. Every stimulus-pattern contained some filled and some unfilled circles. The circle were 0.15 in. in diameter (19.8′ in visual angle) and separated by 0.135-in. spaces (17.8′ in visual angle). The fixation-point coincided with the center of the stimulus-field, *i.e.* between Circles 5 and 6 of the pattern. Thus, with this arrangement fo the circles, the right end of the pattern was projected by each eye to the left cerebral hemisphere, and the left end of the pattern was projected by each eye to the right cerebral hemisphere.

Sometimes the filled elements in the patterns were blackened (producing what are hereafter called black disks) and at other times they were merely grayed (producing what are hereafter called gray disks). Four conditions were thus produced: G/B (filled elements to left of fixation were grayed and those to right of fixation were blackened); B/G (filled elements to left of fixation, blackened; those to right, grayed); B/B (all filled elements blackened); and G/G (all filled elements grayed). The reflectances of the black disks and the gray disks were about 5% and 22%, respectively, while the reflectance of the white paper background was about 85%. These reflectances were estimated by visual comparison to calibrated gray papers.

Both monocular (with each eye) and binocular vision were employed. Two of the 12 *O*s were assigned to each of the six possible sequences of these three conditions. The *O*s were college students with normal vision. They were each first given a binocular practice-session in which the same five stimulus-cards were exposed under the four conditions of intensity. Each *O* served in one 2-hr. session. Each viewing condition always employed each of the 20 patterns once under each element-brightness condition in haphazard order. The inter-trial interval was determined by the pace set by *O*. At no time did *O* know which brightness-condition was to occur on the subsequent exposure.

After each exposure, *O* marked his response on a dittoed score sheet. He merely indicated which circles had been filled in the stimulus-pattern, ignoring whether they were black or gray. An error was counted for an element that was reported incorrectly as filled or as incorrectly not filled. All errors were totalled for the performance-score of each *O*.

Results. The means of errors per element-position are summarized for binocular, right-eye, and left-eye vision in Figs. 2, 3, and 4, respectively. The results are remarkably consistent under these three viewing conditions. In the B/G, B/B, and G/G conditions, fewer errors were made among the elements appearing to the left of fixation. For the G/B condition, there is no marked superiority for the elements on either side of fixation. Under all conditions there is an 'end' effect, *i.e.* the elements at both ends of the pattern are more frequently reported correctly than those in the middle.

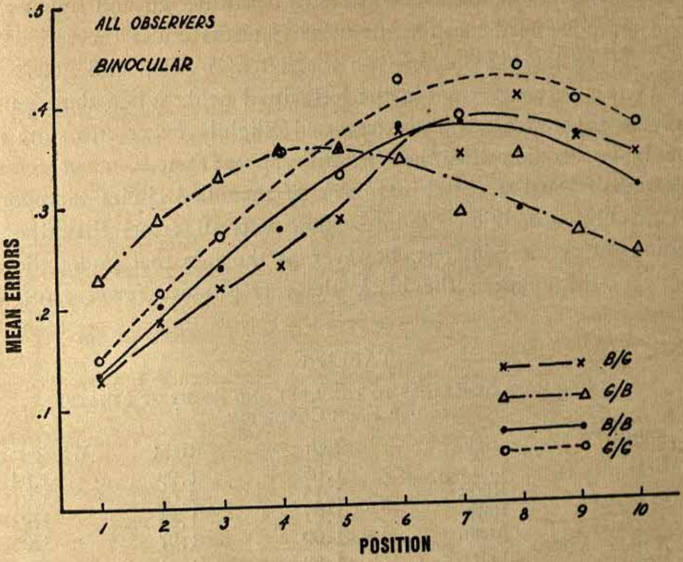


Fig. 2. Means of Erroros per Exposure
(All *O*s: Binocular viewing)

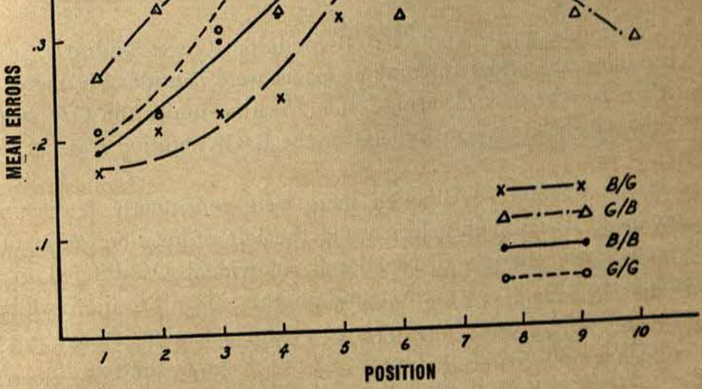


Fig. 3. Means of Errors per Exposure
(*O*s viewing with the right eye)

Table I presents the means of errors made to the left and right of fixation in reproducing the pattern of filled elements under the various conditions. When the black disk is on the left of fixation and gray on the right, there are fewer errors in the left visual field. When these patterns are reversed, the results are also reversed; slightly fewer errors are made to the right of fixation when the black disks appear there.

Clearly, the quality (black or gray) of the filled circles is important, since the G/G condition produces more over-all errors than the B/B condition. Of greater interest, however, is the fact that under the B/G and the G/B conditions the black elements produce fewer errors than the black elements in the control B/B condition. Correspondingly, in the G/B condition the errors for gray elements were considerably more frequent than for the gray elements left of fixation under the G/G condition, although the comparable effect for the B/G condition was not seen.

TABLE I

MEAN NUMBER OF ERRORS TO THE LEFT AND RIGHT OF FIXATION UNDER EACH CONDITION

Condition	Eye	Left	Right	Whole target
B/G	Left	1.05	1.99	3.04
	Right	1.15	2.06	3.21
	Both	1.07	1.85	2.92
	Mean	1.09	1.97	3.06
G/B	Left	1.68	1.41	3.09
	Right	1.70	1.66	3.36
	Both	1.56	1.48	3.04
	Mean	1.65	1.52	3.17
B/B	Left	1.25	1.70	2.95
	Right	1.42	1.91	3.33
	Both	1.23	1.75	2.98
	Mean	1.30	1.79	3.09
G/G	Left	1.56	2.03	3.59
	Right	1.62	2.15	3.77
	Both	1.32	2.03	3.35
	Mean	1.50	2.07	3.57

Individual differences. Although there were consistently fewer errors in reproducing the stimulus-elements to the left of fixation, as in other studies employing similar visual stimuli, individual *O*s differed in the magnitudes of the 'primacy'-effect and in the side of fixation which it favored. A breakdown by *O*s reveals consistent deviations from the average results that are shown in Figs. 2, 3, and 4. From the total of 12 *O*s one may select four who exhibited little difference in over-all accuracy between left and right fields. (Actually, three of these *O*s exhibited slightly fewer errors left of fixation, and the fourth *O* was the only one who made fewer errors for elements on the right of fixation.)

The conclusions to be derived from the data of the selected *O*s are in some respects not the same as those derived from the group of *O*s considered as a whole. For example, the marked superiority for the left-hand elements is not evident. The data of these *O*s for binocular viewing under the B/B and G/G conditions reveal approximately equal numbers of errors right and left of fixation. Therefore, special consideration of possible differences in conclusions for this group of *O*s from the conclusions based on the results of all *O*s will be given in evaluation of the hypotheses tested in this study.

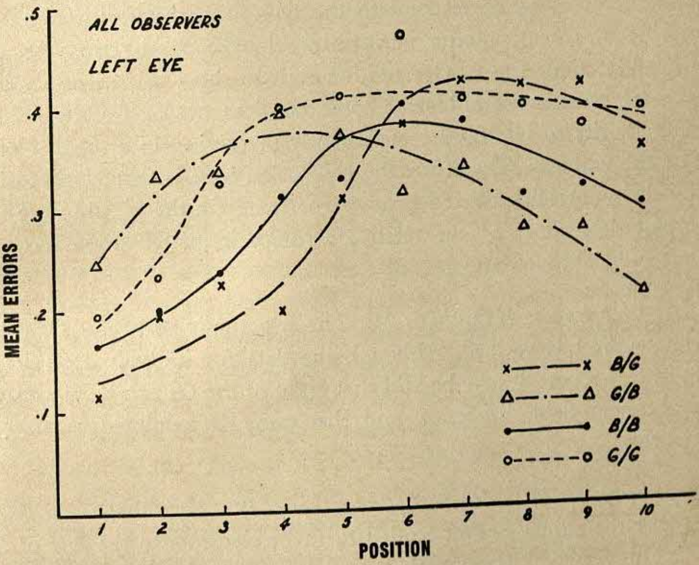


FIG. 4. MEANS OF ERRORS PER EXPOSURE
(*O*s viewing with the left eye)

Accuracy and stimulus-strength:Hypothesis I. Figs. 2, 3, and 4, as well as Table I, reveal a general superiority of the B/B condition over the G/G condition. Therefore, the hypothesis that fewer errors in reproduction-accuracy occur for black stimuli is verified. Inspection of the data for the four selected *O*s indicates, however, little or no difference in total numbers of errors between the B/B and the G/G conditions. We can think of no reason for this.

Effect of viewing eye: Hypothesis II. The evidence seems to corroborate Hypothesis II, which denies any effect of viewing eye, since there were consistently fewer errors for elements to the left of fixation for the results of all *O*s averaged together in each of Figs. 2, 3, and 4. The monocular data of the selected *O*s, however, show them to be atypical with respect to the effect of viewing eye, also. The left eye has, in the B/B condition, a left superiority; but the right eye has a right superiority. The B/G condition shows a greater difference favoring the left field with the left eye than with the right eye. Similarly, the G/B condition reveals a greater differential favoring the right-hand elements with the right eye. These differences all correspond to the typical sensitivity differences for the single eyes, viz. superior performance by the nasal retina (temporal visual field). Only the G/G condition fails to show this correspondence. In general, then, what holds for some *O*s does not necessarily hold for all. A minority of our *O*s are consistently perceiving in a manner unlike that of the majority. Therefore, the conclusion here is that Hypothesis II, as stated, is not substantiated because for *some O*s the viewing eye is an important determiner of the relative accuracy in pattern-reproduction to the left and right of fixation.

Effect of primacy: Hypothesis III. The hypothesis that most *O*s exhibit a primacy-effect is verified. Although the data for individual *O*s are not given in this report, it may be said that eight of the *O*s exhibited a marked left-right hemi-field difference, and that every *O* exhibited a noticeable difference.

Superiority of left elements: Hypothesis IV. Hypothesis IV, which states that the primacy-effect tended to favor elements to the left of fixation, is also verified. Eleven of the 12 *O*s ($p = 0.003$) yielded fewer errors for elements on the left of fixation; and for eight of these *O*s the differences were large.

Stimulus-strength and primacy: Hypothesis V. Hypothesis V states that there is a tendency for primacy to begin at the end of the pattern containing black disks. From Figs. 2, 3, and 4 a large primacy-effect favoring left-hand elements is seen, but this effect is largely nullified by presenting a black disk to the right of fixation, as in the G/B condition. Further, presenting the black disks to the left of fixation produces an even greater effect favoring the left elements than that shown in the G/G and B/B conditions. These effects corroborate the hypothesis. In addition, the four selected *O*s, who did have a below-average amount of left superiority in the G/G and B/B conditions, did produce a right superiority under the G/B condition. Therefore, Hypothesis V appears to be verified.

Discussion. The question as to whether quality or position is the more important determiner of the direction of the primacy-effect now appears to be meaningless unless both the conditions of observation *and* the particular *O* are specified. For example, the tendency for a scan to proceed from black to gray elements might be analogous to a scan from an area of the visual field projecting to the dominant (or non-dominant) cerebral hemisphere toward that hemi-field projecting to the other hemisphere. If this be true, cerebral-hemisphere dominance or possibly other variables must be operating in addition to position.

There is other evidence from earlier studies that the superiority of positions to the left of fixation is due to some attentional variable rather than to sensitivity.[8] Forgays has shown that the superior recognizability of English words exposed tachistoscopically entirely to the right of fixation is directly related to the educational level of the *O*s (children).[9] Thus, this effect is probably the result of learning.

Of course, one cannot assume with finality that darker disks on one side of fixation produced an effect equivalent to hemisphere-dominance. The four atypical *O*s comprising the selected sub-group do not, however, seem to have strong hemisphere-dominance as judged by the approximately equal numbers of errors to the left and right of fixation under the B/B and G/G conditions with binocular viewing. For these *O*s viewing eye is a variable, the effect of which is predictable from sensitivity data. Therefore, the effect of viewing eye is manifested only when there is no marked effect of hemisphere-dominance or some other such factor which could produce left-right differences. This implies that the effect of viewing eye is often masked by a stronger variable which produces consistent superiority for one hemi-field under all viewing conditions.

The equation of fewer errors in one hemi-field with hemisphere-dominance or non-dominance is very likely not entirely valid, of course. Since the modal *O* in the present study performed better for elements to the left of fixation which project to the right cerebral hemisphere, and since most *O*s are conceded to be right-handed (or left-hemisphere dominant), then perhaps stimuli presented to the non-dominant hemisphere are the more "attense," as Dallenbach and his co-workers suggested long ago. Actually, only one of the four *O*s who show the left-superiority to the greatest degree is left-handed, but three of the four *O*s in the so-called symmetrical (atypical) group are left-handed. There may, however, be some other unknown factor operating which is confounded with, or difficult to discriminate from, hemisphere-dominance. There is some evidence for such a factor in various studies by Anderson[10], Mishkin and Forgays,[11] Orbach,[12] and Heron.[13] They have found that performance to the left or right of fixation is not entirely stable for an *O* under all conditions of observation and stimulus-material, as would surely be expected if hemisphere-dominance were the only variable. Hence, some other factor or factors favoring certain locations or selected parts of familiar stimulus-complexes must be present. Such a factor could be the hypothetical scanning mechanism proposed earlier. Since the directional sequence of this scan is presumably established through experience, it should be one that is more effective in perception of the specific visual pattern. For many visual tasks, such as reading English, the conventional and well-practiced scan proceeds from left to right, because written English has a built-in directionality. In addition to the left-to-right sequence of words and letters, a large number of English words have the root, which is the primary carrier of meaning, at the beginning of the word. Hence, a learned, 'scanning set' could be the unknown factor.

---

[8] S. B. Anderson and Sherman Ross, Memory for items in a matrix, this JOURNAL, 68, 1955, 595-604; H. R. Brandt, Ocular patterns in visual learning, this JOURNAL, 54, 1941, 530-535.

[9] D. G. Forgays, The development of differential word-recognition, *J. exp. Psychol.*, 45, 1953, 165-168.

It should be pointed out that in this experiment the stimuli were 'meaningless' patterns of blackened and open circles, with each element radially symmetrical. Investigation of the basic perceptual mechanism using these non-directional stimuli was not therefore, contaminated by the meaningfulness, dependencies, or other discriminating characteristics usually associated with printed words. Nevertheless, the left-hand elements were found to be favored by the alleged primacy-effect. The direction of the hypothetical scan must, however, vary among *O*s.

A few words about the implications of the present conclusions for an understanding of reading disability are in order. First, in the mature reader the left-to-right direction of reading may be assumed to have been thoroughly established. This left-to-right set apparently takes precedence over all other factors such as eye-dominance, 'mixed' dominance, etc. It is proposed that the difficulty of learning to read is directly related to the difficulty in establishing the left-to-right sequence in perceiving. Certain situations of eye- or hemisphere-dominance could predispose the child to scan the visual field from right to left. These 'incorrect' scans may be more effective for the pre-reading visual experiences of the particular child, and, therefore, lead to the establishment of a right-to-left set. If this produces negative transfer to the learning-to-read task, then a reading disability—particularly strephosymbolia—might ensue. Since this situation is due to the learning of incorrect viewing sequences, then learning-principles can be employed in unlearning them. If other factors such as emotional concomitants do not arise, then the learning of the proper set will not be unduly delayed.

---

[10] Anderson, *op. cit.,* 37-40.

[11] Mortimer Mishkin, and D. G. Forgays, Word-recognition as a function of retinal locus, *J. exp. Psychol.,* 43, 1952, 43-48.

[12] Jacob Orbach, Retinal locus as a factor in the recognition of visually perceived words, this JOURNAL, 65, 1952, 555-562.

[13] Heron, *op. cit.,* 46.

Summary. The purpose of this study is to investigate possible variables affecting the relative accuracy with which *O*s reproduce complex stimulus-patterns presented in the left and right halves of the visual field. Test-patterns, formed by variously darkening 5 circles in a 10-circle template, were tachistoscopically exposed across fixation. Eye-dominance was simulated for *O* by monocular viewing, and hemisphere-dominance was simulated for *O* by varying differentially the contrast of the darkened stimulus-elements which appeared in the right or left visual field.

The conclusion based on the data of all *O*s indicates superior accuracy for the left-field elements regardless of viewing eye. For the *O*s who do not, however, exhibit the general left-superiority, an effect of viewing eye is seen. This effect is consistent with the superior sensitivity of the nasal retina. Left-right field differences are not uniquely related to the handedness of the *O*s.

These data suggest that there is a non-structural and, therefore, presumably learned tendency favoring for most adult *O*s the elements to the left of fixation in complex patterns such as the ones used in this study. This tendency supersedes all effects of eye or hemisphere dominance, which may be important if the learned left-superiority has not been strongly established.

MEDIATED GENERALIZATION AS A FUNCTION OF SEMANTIC DIFFERENTIAL PERFORMANCE

By JAMES C. BAXTER, University of Texas

This study was designed to test two implications of Osgood's theory of meaning.[1] According to this position, the meaning of a sign is contained in the manifold of representational, mediative reactions associated with it and hence is aroused in the subject (*S*) upon its subsequent presentation. The mediators associated with most signs are thought to be multidimensional, as a result of diverse associative experiences; and the pairs of polar reactions defining each of the several dimensions are presumed to be reciprocally antagonistic in function.[2] At the level of linguistic behavior, Osgood has proposed the semantic differential as a measure relevant to these processes. This instrument allows *S*s to describe the meaning of a variety of signs by means of patterns of scaled judgments made to pairs of contrasting adjectives. Factor analyses of these scales across a broad range of signs have indicated that their structure can be meaningfully described by a relatively small number of orthogonal coördinates in a semantic space.

In an effort to draw this position into closer alignment, Osgood has recently proposed additional relationships between theory and test. He has assumed that "there is some finite number of representational mediation reactions available to the organism, [and] that the number of these alternative reactions . . . corresponds to the number of dimensions or factors in semantic space. Corresponding to each of the major dimensions of the semantic space, defined by a pair of polar terms, is a pair of reciprocally antagonistic mediating reactions, which we may symbolize as r_{MI} and $\bar{r}_{MI}$. for the first dimension, r_{MII} and $\bar{r}_{MII}$ for the second dimension, and so forth."[3] Accordingly, point locations in semantic space are taken to reflect the composition and intensity of the mediators aroused, and multidimensional distance between points is adopted as a measure of meaning similarity.

Although experimental verification of this relationship between theory and test would seem to be of central importance to Osgood's position, a recent review of

---

* Received for publication November 17, 1960. This paper is based on a dissertation submitted to the Graduate School of The University of Texas in partial fulfillment of the requirements for the Ph.D. degree.

[1] C. E. Osgood, The nature and measurement of meaning, *Psychol. Bull.*, 49, 1952, 197-237; *Method and Theory in Experimental Psychology*, 1953, 690-699. C. E. Osgood, G. J. Suci, and P. H. Tannenbaum, *The Measurement of Meaning*, 1957, 5-30.

[2] Osgood, Meaningful similarity and interference in learning, *J. exp. Psychol.*, 36, 1946, 277-301; An investigation into the causes of retroactive interference, *J. exp. Psychol.*, 38, 1948, 132-154; The similarity paradox in human learning, *Psychol. Rev.*, 56, 1949, 132-143.

[3] Osgood, Suci, and Tannenbaum, *op. cit.*, 1957, 27.

the literature has revealed only two lines of evidence presently available. The first, representing several studies demonstrating conditioning and interaction-effects between mediators, is somewhat indirect;[4] whereas the second, a series of studies investigating mediated generalization, is of doubtful confirmatory value.[5] Although Dicken was able to show significant differences in generalization between clusters of words varying in semantic profiles, he also found anomalous differences in generalization between theoretically equivalent words within clusters. Moreover, Staats *et al.* were unable to demonstrate generalization between signs as a function of semantic similarity.

The present study employed a paradigm of mediated generalization in which the galvanic skin-response (*GSR*) was conditioned to one of a sample of experimental signs (words) by means of contiguous pairing with a noxious significate (electric shock). Under these conditions, two hypotheses were advanced: (I) It was expected that the amplitude of the *GSR* generalizing to noncritcial test-words should bear an inverse relationship to the multidimensional distance (*D*) in semantic space between these words and the critical word employed in response-acquisition; (II) The second hypothesis was aimed at evaluating Osgood's principle of reciprocal inhibition. Since noxious stimulation was used as a significate during training, it was expected that the mediational part (r_M) of the observable behavior (*GSR*) associated with this stimulation should correspond primarily to the negative pole of the evaluative dimension. If Osgood's position is sound, acquisition of the *GSR* should be facilitated in association with negatively evaluated words, whereas generalization to positively evaluated words should be antagonized. On the other hand, acquisition of the response in association with positive words should be antagonized, with facilitated generalization to negative words. Differences in the slopes of generalization gradients were expected to occur, then, as a function of the evaluative meaning of the critical word used during training.

METHOD

Semantic differential. A sample of 6 words was selected from the atlas of 360 words whose semantic differential profiles have been described by Jenkins.[6] Words were chosen on the basis of four specific criteria: (1) diverse representation on the evaluative dimension; (2) lack of readily detectable physical similarity; (3) absence of obvious denotative similarity; and (4) lack of obvious, disorganizing emotional connotations. The words selected were *home, nurse, money, bleak, fraud,* and *starve.* Evaluative weightings (estimated from *good-bad*) were 6.53, 6.40, 5.63, 2.73, 1.63, and 1.50 respectively; activity weightings (estimated from *active-passive*) were 5.23, 6.07, 5.80, 2.83, 4.97, and 4.10 respectively; and potency weightings (estimated by *hard-soft*) were 2.77, 2.43, 5.87, 4.53, 5:67, and 6.53 respectively.[7]

---

[4] D. C. Ferguson, Unpublished manuscript, University of Illinois, 1956; C. K. Staats and A. W. Staats, Meaning established by classical conditioning, *J. exp. Psychol.,* 54, 1957, 74-80; Attitudes established by classical conditioning, *J. abn. soc. Psychol.,* 57, 1958, 37-40; A. W. Staats, C. K. Staats, and D. A. Biggs, Meaning of verbal stimuli changed by conditioning, this JOURNAL, 71, 1958, 429-431; A. W. Staats, C. K. Staats, W. G. Heard, and L. P. Nims, Replication report: meaning established by classical conditioning, *J. exp. Psychol.,* 57, 1959, 64.

[5] C. F. Dicken, Connotative meaning as a determinant of stimulus generalization, Unpublished doctoral dissertation, University of Illinois, 1958; A. W. Staats, C. K. Staats, and H. L. Crawford, First-order conditioning of meaning and the parallel conditioning of a G.S.R., Unpublished manuscript, Arizona State College, 1958.

Twelve adjectival scales were chosen, with four scales representing each of the three major factors in semantic space. Criteria for selection were: (a) high, relatively specific loadings on the dimension of interest; and (b) relatively direct application to the experimental words. These scales and the six words were presented in a graphic form of the semantic differential. Three different test-forms were devised, each with scale-order and adjective-position randomized. Instructions were adapted from Osgood.[8]

Apparatus. An exposure apparatus, developed in the Texas laboratory, was used to present the experimental words to *S*.[9] By means of this apparatus the stimulus-cards could be exposed consecutively in any desired order and for variable lengths of time. The experimental words were printed in India ink on white cards in letters approximately ¼-in. high. Blank filler cards were also prepared, which were similar to the others in every respect except that no words were printed on them.

A Fels dermohmmeter, connected to an Easterline Angus graphic recorder, provided a continuous record of *S*'s *GSR*. Electrodes from the dermohmmeter were attached to *S*'s palms, and were taped on to insure uniform contact. A microswitch on the back of the exposure-shutter activated a Hunter timer, which in turn activated a Bird inductorium. This instrument, connected to *S*'s left wrist, constituted the shocking circuit. The timer also signaled the occurrence of exposures on the recorder.

Experimental series. Twenty separate cards were made for every word, along with five blank cards (total = 125). In this way it was possible to 'stack' the entire series for an individual *S*. Word-orders were prearranged for every *S*, except for the choice of his critical word. These series were arranged by the following plan (randomized for each *S*). At the outset of the series, a succession of 18 pre-training trials was given in order to familiarize *S* with the apparatus and to extinguish existing *GSR*s. This block of trials consisted of three presentations of each of the six words, arranged in random orders within blocks of six words. Immediately following the pre-training trials, a block of 17 training trials was presented, consisting of 12 reinforced (shock) exposures of the critical word, along with five non-reinforced, randomly interspersed blank cards (a constant order for blank cards was used for all *Ss*). The blank cards were employed in an effort to extinguish respones to cues from the apparatus. A block of generalization-trials immediately followed the training block. This block was composed of one randomly ordered presentation of each of the non-critical words. Three non-reinforced repetitions of the critical word followed the generalization-block, providing data for the presence of conditioning.

---

[6] J. J. Jenkins, W A.. Russell, and G. J. Suci, An atlas of semantic profiles for 360 words, this JOURNAL, 71, 1958, 688-699.

[7] Judgments inclining toward 7.00 represent the positive, active, or potent poles of the three dimensions (from Jenkins *et al.*).

[8] Osgood, Suci, and Tannenbaum, *op. cit.*, 1957, 82.

[9] K. M. Dallenbach, A simple and inexpensive card-changer, this JOURNAL, 73, 1960, 139-141.

Subjects and procedure. The *Ss* (60 men and women), from undergraduate courses in psychology, were randomly subdivided into 3 groups of 20 each.

Individual sessions were conducted with each *S* according to the following plan. Upon entering the experimental room, *S* was seated in a comfortable chair in front of the apparatus and given general instructions to the task. Stress was placed upon the task as a study of the relationship between thought and physiological processes. Following the standardized introductory set, *S* was presented the semantic differential with attached instructions, and asked to complete it. Any questions raised were answered in general terms. After *S* had begun working, the experimenter (*E*) casually checked to insure that instructions were being followed appropriately. A few *Ss* failed to understand, and required additional instructions at this point.

Following completion of the semantic differential, the *GSR*- and shock-electrodes were attached to *S*'s palms and wrist respectively. He was asked to assume a comfortable position and to hold it with arms relaxed and his fingers resting on the table in front of him. He was instructed to focus his eyes on the shutter of the apparatus, where a series of words would be presented. He was further instructed to pronounce each word subvocally when it was exposed but to keep his mind as blank as possible between exposures. After this, the inductorium was adjusted to *S* to a degree that the shock was definitely unpleasant but not painful. He was then given a 3-min. rest before the experimental series was begun.

During this interval, the evaluative dimension of *S*'s semantic differential was key-scored, and a critical word was assigned according to *S*'s group-membership. *Ss* of Group I were assigned one of the six experimental words at random; whereas members of Groups II and III were assigned words judged extreme on the evaluative dimension. The critical word for the *Ss* of Group II was the one he judged most positive. The *Ss* of Group III, on the other hand, received the word they had judged to be most negative. Choice was random in cases of ties.

Following the choice of *S*'s critical word, the experimental words were organized in their predetermined order and placed in the apparatus. At the completion of the rest-interval, the dermohmmeter was calibrated, and the experimental series was begun. Exposure-time for each word was approximately 2 sec., with an inter-trial interval of approximately 20 sec. Shock was associated with the critical word only during the training phase of the experimental series. The shock occurred 0.5 sec. after the appearance of the word and continued until the word was obscured by the shutter.

Results. Logarithmic transformations of micromho conductance units were used since these measures seem to approximate best the assumptions of parametric analysis.[10] Conductance change was determined by subtracting the reciprocal of *S*'s resistance at the point of presentation from the reciprocal of his resistance at the point of maximal deflection occurring within an interval of 5 sec. after presentation. These measures were then converted to micromho conductance units, *i.e.* conductance $\times 10^6$, a constant was added, and log values were determined from a four-place table.

---

[10] E. A. Haggard, On the application of analysis of variance to G.S.R. data: I. The selection of an appropriate measure, *J. exp. Psychol.*, 39, 1949, 378-392; III. Some effects of the use of inappropriate measures, *ibid.*, 39, 1949, 861-867.

Training. The present design permitted evaluation of the adequacy of training on two counts. First, by comparing the mean *GSR* deflection of the three pretraining exposures of the critical word with that of the three post-training exposures, a measure of conditioning was derived. Further, an index of differential conditioning was provided by comparing the mean *GSR* deflection of the last three training exposures of the critical word with that of the last three blank card exposures.

TABLE I

MEAN DIFFERENCES IN *GSR* BETWEEN PRE- AND POST-TRAINING AND BETWEEN CARDS WITH CRITICAL WORDS, AND THOSE THAT ARE BLANK
(N=20 for each group)

Comparison	Group	Mean diff.	*SE*	*t*
Pre- vs. post-training	I	.238	.024	9.90*
	II	.246	.033	7.45*
	III	.129	.042	3.07*
Critical word vs. blank card	I	.196	.024	8.21*
	II	.224	.029	7.69*
	III	.100	.025	4.00*

* Significant at or beyond the 5% point (one-tailed).

The results of these two comparisons are summarized in Table I. Response-levels between pre- and post-training exposures of the critical word and between critical word and blank card exposures were significantly different in all three groups ($p < 0.05$).[11] Accordingly, it was assumed that an acceptable degree of differential conditioning was achieved in each group during training.

Generalization. Hypothesis I was that the amplitude of *GSR* generalization to non-critical test-words should bear an inverse relationship to their multidimensional distance from the critical word. Accordingly, *Ss* of Group I were assigned critical words at random from among the six words. *GSR* results for the remaining five words constituted generalized responses.

Difference-measures were found for each generalization-word for every *S* by subtracting its response-amplitude on the last pre-training trial from that of its test, or generalization-trial; hence, variation in the amplitude of the response to these words was independent of S's habituation to the situation, and presumably a function of the training itself. The test-words were then ranked in terms of their multidimensional distance from the critical word for each individual *S*. Corresponding measures of response were then entered in a treatment by *S*-analysis of variance table, with the five treatments defined as increasing multidimensional distances from the critical word. It was expected that treatment-means would reflect a decreasing monotonic trend as a function of multidimensional distance. An analysis of the variance of these means failed to reach significance at the 5% level of confidence ($F < 1.00$). The comparison indicated a lack of systematic variation in response-means as a function of multidimensional distance, and thus failed to support the hypothesis.

---

[11] The 5% level of confidence (one-tailed) is accepted as significant throughout this report.

Hypothesis II predicted interaction of generalization-gradients as a function of the location of the critical word on the evaluative dimension of the semantic differential. Accordingly, *S*s in Group II were assigned the words they judged most positive, while those in Group III were assigned the words they judged negative. Generalization-words were adjusted for pre-training level in the same manner as was done for the *S*s of Group I. The distance of each word was then ranked from the critical word (evaluative dimension only), and corresponding *GSR*-values were entered in an analysis of variance table. The design employed to test this hypothesis was that which Lindquist designates a Type I trend analysis design.[12] Interaction of the several means (unidimensional distance) between groups was of principal interest by way of reflecting an interaction in generalization trends.

Before conducting the analysis for over-all trend, it was necessary to consider first the comparability of the group-trends involved. The problem involved demonstrating the comparability of distance-intervals between groups. A test of this condition was provided by computing the median distance associated with each treatment (across both groups). Median tests were then conducted for each distance treatment in turn, to determine whether distances within both groups could be regarded as having come from the same populations, each possessing these over-all medians. In no case did the Chi-square reach significance at the 5% level of confidence. It was concluded, therefore, that the distances could be regarded as comparable between groups. These medians were adopted as intervals for both groups.

---

[12] E. F. Lindquist, *Design and Analysis of Experiments in Psychology and Education*, 1953, 267-273.

A summary analysis of the generalization data of Hypothesis II is presented in Table II. This analysis indicates that the comparison of principal interest, *i.e.* trend-interaction, reached statistical significance at the 5% level of confidence. It is also evident from the table that the over-all response-level between groups (trend-level) does not differ significantly ($p > 0.05$). The mean *GSR*-amplitude for each evaluative distance is presented graphically in Fig. 1. It can be seen that the two trends did interact as expected, and that an intersection of the gradients occurred somewhere between Distance-points 2 and 3.

TABLE II

ANALYSIS OF THE VARIANCE OF *GSR* GENERALIZATION AS A FUNCTION OF UNIDIMENSIONAL DISTANCE (Groups II and III)

Source	*df*	Mean square	*F*
Independent			
Subjects	39	.091	
Groups	1	.004	.05
Error	38	.094	
Dependent			
Within Subjects	160	.024	
Distance	4	.021	
Distance×Group (Trend interaction)	4	.058	2.52*
Error	152	.023	
Total	199		

* Significant at the 5% point.

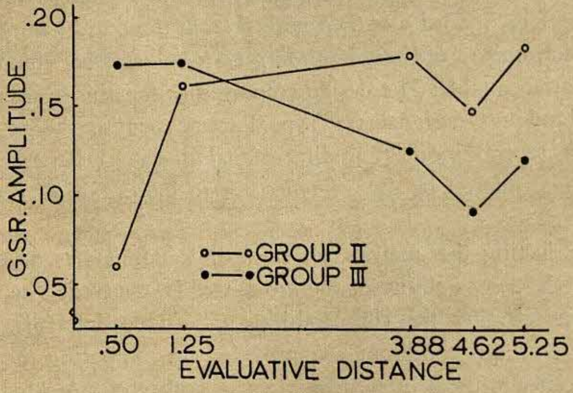


FIG. 1. THE EFFECT OF EVALUATING MEANING OF THE CRITICAL WORD ON *GSR* GENERALIZATION

Following completion of this analysis, differences in response-means between groups were computed to evaluate their individual significance. These differences were found to be 0.11, 0.01, −0.05, −0.06, and −0.07 for treatments 1 through 5 respectively. Tukey's Wholly Significant Difference Test (*WSD*) was employed in evaluating the several differences.[13] Differences associated with Treatments 1 and 5 exceeded the *WSD*-value at the 5% level (*WSD* = 0.06). It was, therefore, concluded that the present data confirmed Hypothesis II.

Replication. It was thought advisable to replicate the test of Hypothesis I to assess the reliability of the result found in the main experiment. In addition, the new group of *Ss* also provided an opportunity to evaluate the predicted relationship with an alternative response which was simultaneously recorded. Since electric shock was again employed as the significate during training, it was possible to record the mild tonic flexion response which reliably accompanies this stimulation, and then observe its generalization in addition to that of the *GSR*.

TABLE III

MEAN DIFFERENCES IN RESPONSE AMPLITUDE BETWEEN PRE- AND POST-TRAINING AND BETWEEN CARDS WITH CRITICAL WORDS AND THOSE THAT ARE BLANK

Response	Comparison	Mean diff.	*SE*	*t*
GSR (N=20)	Pre- vs. Post-Training	.228	.032	7.14*
	Critical Word-Blank Card	.253	.027	9.36*
Tonic flexion (N=14)	Pre- vs. Post-Training	4.53	2.31	1.96*
	Critical Word-Blank Card	15.46	3.30	4.70*

* Significant at or beyond the 5% point (one-tailed).

To record this flexion-response, a stabilometer was devised by making appropriate modifications on a standard Keeler polygraph. The stabilometer was fashioned by enclosing the rubber bulb of the sphygmomanometer (inflated to 10 mm. Hg) in a small wooden container with a 'floating' top. Thus, slight variations in pressure on this bulb were continuously recorded on the polygraph. Word-exposures were also recorded on this apparatus by triggering the pneumograph harness with the shutter of the exposure-apparatus. This tonic flexion-response was recorded simultaneously with the *GSR*.

Twenty men and women *Ss* (Group IV) were recruited from introductory courses in psychology. Individual experimental sessions were conducted according to the same procedure used with Group I, with one exception; the *Ss* were asked to rest their hands on the 'platform' on the table, rather than on the table itself.

GSR-responses were recorded, transformed, and analyzed by the method described above. Our prediction was the same as in Hypothesis I of the main experiment. The results of the differential training, *i.e.* differences in amplitude of the responses to critical words between pre- and post-training and between cards with critical words and those that are blank, are presented in Table III. Both of these differences were reliable at the 5% point of confidence. It was concluded, therefore, that satisfactory differential training had again occurred. An analysis of the trend of the means of the responses as a function of multidimensional distance was conducted as explained above. The variance of these means failed to reach significance ($F < 1.00$).

---

[13] T. A. Ryan, Multiple comparisons in psychological research, *Psychol. Bull.*, 56, 1959, 26-47.

The data obtained from the stabilometer were analyzed by a procedure analogous to that used with the *GSR*-data. Tonic flexion-response was defined as the maximal stabilometer deviation, in either recorded direction, which occurred within a period of 5 sec. after the stimulus-word was exposed. Deviations were measured to the nearest 0.25 mm. Six *S*s were excluded from this group as a result of instrument failure, reducing the sample to 14.

The results of differential training are presented in Table III. These comparisons were reliable at the 5% level of confidence. Results of generalization were determined by subtracting the pre-training levels of response for each test-word from its test-level for each *S*, and entering these values in a table of analysis of variance. It was expected that these means would reflect a decreasing trend. The variability in these means did not, however, reach statistical significance at the 5% level of confidence ($F < 1.00$). The hypothesis also failed to find support with this alternate method of measuring responses.

Discussion. The results of this study provide only limited support for the position of Osgood *et al.* While confirmation of the predicted interaction of generalization-gradients is congruent with Osgood's proposed antagonistic relationship between mediators of the evaluative dimension, the failure of multidimensional distance to predict successfully the order of magnitude of a generalizing response is in contradiction. Further, replication of this test, using both the *GSR* and a simultaneously recorded tonic flexion-response, provided comparable results. In view of these data, it seems clear that multidimensional distance is unable to provide a measure of similarity capable of predicting generalization under these conditions.

Several questions are raised by the results of the experiments testing Hypothesis I. Since a relatively small sample of signs was used, which was selected on the basis of broad evaluative representation, it could be argued that adequate sampling of the semantic space was not achieved. Repeated evidence of the orthogonality of these dimensions[14] suggests, however, that this procedure should not drastically interfere with representation on other dimensions. Furthermore, data from Jenkins substantiate this expectation with respect to these signs on the three dimensions used.[15] A related possibility is that the signs used were representative of some unknown class of signs whose semantic composition is poorly reflected by the differential. While theoretical statements relative to this question are lacking,

---

[14] Osgood, Suci, and Tannenbaum, *op. cit.*, 31-75.

[15] Jenkins, Russell, and Suci, *op. cit.*, 688-699.

Osgood has identified some signs which seem to possess special properties, viz. political candidates and election issues.[16] Similarly, Dicken has found that semantic differential-defined synonymity corresponds poorly to a mediated generalization definition for some signs.[17]

These data also raise the question of the class of responses which should be expected to provide mediators corresponding to the semantic dimensions. While this study assumed that the mediators associated with the significate behavior (*GSR*) elicited by shock should correspond *primarily* to the evaluative dimension, no assumption of its *unique* correspondence to this dimension is warranted. Certainly mediators from this behavior as well as those from other associated behaviors (*e.g.* tonic flexion) could be expected to relate to a number of dimensions. The present data do not, however, clarify this point.

Multidimensional distance (*D*) as an index of similarity may also possess some characteristics of significance to this study. One property of this measure is that of weighting equally each of the factors contributing to the measure. Since previous evidence has shown that the evaluative dimension accounts for approximately twice the scale variance attributable to the other two major dimensions, it is possible that some bias is inherent in the index.[18] For example, Osgood has speculated that evaluative meaning may be of more "cognitive importance" than the other dimensions presently known. If this should be the case, it is reasonable to expect that differential weighting of this dimension would improve prediction of semantic similarity. While these questions may be raised by the results of this study, no data are available which bear upon their answer.

SUMMARY

The order of magnitude of *GSR*-generalization among a sample of six words was predicted on the basis of semantic differential performance. It was hypothesized that (I) the amplitude of the *GSR* generalizing to non-critical test-words would bear an inverse relationship to the multidimensional distance of these words from the critical word; and (II) differences in the slopes of generalization would occur as a function of the evaluative meaning of the critical word. Sixty men and women were assigned to three groups of 20 *S*s each. Critical words were randomly chosen for Group I, while the *S*s of Groups II and III were assigned words judged

---

[16] Osgood, Suci, and Tannenbaum, *op. cit.*, 1957, 104-124.

[17] Dicken, *op. cit.*

[18] Osgood and Suci, Factor analysis of meaning, *J. exp. Psychol.*, 50, 1950, 325-338; Osgood, Suci, and Tannenbaum, *op. cit.*, 31-75.

to be extreme on the evaluative dimension. Generalization of the response to non-critical words was analyzed.

Hypothesis I was not supported in the present study. No orderly variation in response-amplitude was found as a function of multidimensional distance. Replication of this part of the study, using both the *GSR* and a tonic flexion response, also failed to support the hypothesis. Hypothesis II was supported by the data ($p < 0.05$). The slopes of generalization between words interacted as predicted.

Results of the study were interpreted as offering limited support for Osgood's position. Support of Hypothesis II was taken as confirmation of Osgood's principle of reciprocal inhibition with respect to the evaluative dimension. The data obtained in testing Hypothesis I were interpreted as indicating a failure of multidimensional distance to provide an effective measure of meaning similarity. Some possible reasons for this failure were discussed.

EXPECTATION AND RESISTANCE TO EXTINCTION UNDER PARTIAL REINFORCEMENT AND RISK-TAKING

By DONALD J. LEWIS, Rutgers University, and CARL P. DUNCAN
Northwestern University

Like most studies of extinction after partial reinforcement, our earlier experiments have not required *S* to risk anything but his own effort; *S* either won or did not win, but did not lose, money each time he responded.[1] The present experiment introduced some risk-taking, that is, *S* could lose money. He was given an initial stake, the amount of which, he was led to believe, he could either add to or decrease each time he played the apparatus. The purpose of the experiment was to determine what effect risk-taking would have on resistance to extinction, particularly whether risk-taking would grossly modify the relation, found in earlier studies, of resistance to extinction and percentage of reinforcement.[2]

Method: (1) *Apparatus.* The apparatus has been described in detail elsewhere;[3] it is an electrical slot machine which can be prearranged to pay off on specified trials or plays. *S* inserts a chromium-plated disk into an aperture, pushes one button in each of four columns of four buttons, and pulls a lever. The machine makes 'business-like' noises, at the conclusion of which a disk is dropped into the pay-off tray if it is a pay-off trial.

(2) Design. Five percentages of reinforcement, 0, 11, 33, 67 and 100, were combined factorially with two magnitudes of reward, 1¢ and 5¢, making ten conditions in all. Separate groups of 30 *S*s each were run under every condition. For all groups, the first nine plays made up the trials of 'acquisition.' During acquisition, the 0%-Group received no pay-offs; the 11%-Group was paid off only on Play 9; the 33%-Group was paid off on Plays 2, 5, and 9; the 67%-Group on Plays 1, 2, 4, 5, 6,

---

* Received for publication June 29, 1959. This study was supported by a grant from the National Science Foundation. The authors thank Miss Barbara Rennie for her assistance.

[1] For a review of earlier studies see D. J. Lewis, Partial reinforcement since 1950: A selective review of the literature, *Psychol. Bull.*, 57, 1960, 1-18. The studies by the present authors include: D. J. Lewis and C. P. Duncan, Effect of different percentages of money reward on extinction of a lever-pulling response, *J. exp. Psychol.*, 52, 1956, 23-27; Expectation and resistance to extinction of a lever-pulling response as functions of percentage of reinforcement and amount of reward, *ibid.*, 54, 1957, 115-120; Expectation and resistance to extinction of a level-pulling response as a function of percentage of reinforcement and number of acquisition trials, *ibid.*, 55, 1958, 111-128; Vicarious experience and partial reinforcement, *J. abnorm. soc. Psychol.*, 57, 1958, 321-326.

[2] See particularly Lewis and Duncan, *opp. citt.*, 1956 and 1957.

[3] Lewis and Duncan, *op. cit.*, 1957, 115.

and 9; and the 100%-Group on all of the plays. After the ninth play, none of the *S*s ever won again.

A record was kept of the total number of plays, and, in addition, the *S*s of all Groups were asked to state their expectations of winning or not winning on every play. Secured to the front of the machine was a rating scale with the numbers 1 through 6 placed at equal intervals above a line. Above these numbers was the heading, "Expectation of winning or not winning." Corresponding to the numerical scale-points were the descriptions: "fairly sure of not winning," "mildly sure of not winning," "slightly sure of not winning," "slightly sure of winning," "mildly sure of winning," and "fairly sure of winning." On each play, before he pulled the lever, *S* called out the numerical value corresponding to his level of expectancy; these numbers were recorded by *E*.

(3) Instructions. First, *S* was told briefly how to operate the machine. He was to place a disk in the slot, push one button in each of the four columns of buttons, and then pull the lever. If he had picked the right combination of buttons, he was informed, the machine would pay off a disk. He was asked to state his expectation before each play. Then he was given the following instructions:

> Now each of these disks is worth 1¢ [or 5¢] and to start with, I shall stake you to 60 disks, so you have 60¢ [or $3.00] to begin with. This is your stake in this box [*E* indicates small box containing 60 disks at right side of apparatus.] You play the machine with disks from your stake. If the machine pays off—it pays off only one disk at a time—you will get back the disk you put in the machine, and in addition I shall give you a disk so you will have won 1¢ [or 5¢]. If the machine does not pay off, you will have lost the disk you put in the machine, so you will have lost 1¢ [or 5¢]. Thus you may win a lot more than you have to start with, or you may lose all you have. You can play as long as you wish. When you decide to quit, I shall give you 1¢ [or 5¢] for every disk you have in the box.

The *E* dropped a disk in the box each time *S* won. After *S* had decided to quit, he was paid off with money, if he had any disks left; given a plausible but false explanation of the machine; and told not to tell any other person how the machine operated or how much he had won.

(4) Subjects. The *S*s, 300 in number, were drawn from introductory courses in psychology at Northwestern University and were assigned to the various experimental groups in turn as they appeared in the laboratory. Thirty *S*s were assigned to each of the 10 groups.

Since the maximal number of plays that could be made by the *S*s in the 0%-Group was 60 (the number of disks in the initial stake), all the *S*s in any group who played through all their disks, including any won, were assigned a score of 60.

Results: (1) Number of plays. If partial reinforcement operated to produce greater resistance to extinction, then the number of *S*s making the maximal number of plays (60) should be greater for the groups with the lower percentages of reinforcement than for groups with the higher percentages. To test for the presence of such a trend, a 2 × 2 table was constructed. The *S*s (150) receiving the least reinforcement (those in the 0%, 11%, and half of those in the 33% Groups) and the *S*s (150) receiving the most reinforcement (half of those in the 33%, and all those in the 67% and 100% groups) were divided into two groups: those who played the maximal number of times (60) and those who stopped short of the maximum. Of the least-reinforced groups, 68.5 played the maximal number of times and 81.5 stopped short of the maximum. Of the more reinforced groups, 47.5 played the maximal number of times and 102.5 played less than that number. These frequencies yielded $\chi^2 = 6.198$, which is significant at the 2% level.

In another analysis, the number of plays made by each *S* was converted to a common logarithm (to permit comparison with previous studies) and an analysis of variance was performed on these transformed scores. Reinforcement was significant at the 5% level ($F = 2.60$, 4 and 290 *df.*). Neither amount of reinforcement ($F = 2.67$, 1 and 290 *df.*), nor the interaction of amount and percentage ($F = 1.80$, 4 and 290 *df.*) was significant. Accordingly the results for the 30 *S*s rewarded with 1¢ and the 30 rewarded with 5¢ were combined at each value of percentage of reinforcement and are plotted as the lower curve shown in Fig. 1. In spite of some irregularities in the curve, there was, in general, less resistance to extinction on the part of groups with higher percentages of reinforcement.

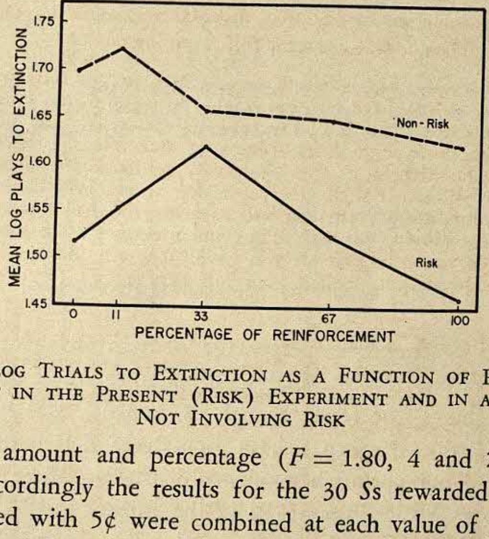


FIG. 1. MEAN LOG TRIALS TO EXTINCTION AS A FUNCTION OF PERCENTAGE OF REINFORCEMENT IN THE PRESENT (RISK) EXPERIMENT AND IN AN EXPERIMENT NOT INVOLVING RISK

The chief purpose of the present experiment was to determine if the element of risk, that is, the necessity of using one's stock of valuable tokens, produced any significant change in the function relating resistance to extinction and percentage of reinforcement. As a comparison with the present results, some data taken from an earlier study are plotted as the upper curve in Fig 1.[4] The *S*s represented in the upper curve were run under the same five percentages of reinforcement as were used in the present experiment, but without the element of risk, *i.e.* the *S*s either won or did not win, but did not lose, on each of the nine trials of acquisition. Of the 20 *S*s represented at each of the five points in the upper curve, half had been rewarded with 1¢, half with 10¢, for each disk they won; these subgroups were combined, since resistance to extinction did not differ significantly as a function of these two amounts of reward. As with the *S*s of the present study, any *S* from the previous study who had made more than 60 plays was assigned a score of 60.

---

[4] Lewis and Duncan, *op. cit.*, 1957, 116. Note that only the groups from the 1957 study that were rewarded with 1¢ or 10¢ are used for comparison here.

The results of the two experiments, depicted in Fig. 1, were compared by an analysis of variance which is summarized in Table I. The first variable, risk vs. non-risk, represents the two experiments. As usual, the *F* for percentage of reinforcement was significant; in general, resistance to extinction decreased as percentage of reinforcement increased. The other two *F*-values from the analysis of variance are of more interest here. As Table I shows, the term for risk vs. non-risk was highly significant, but the interaction between risk vs. non-risk and the percentage of reinforcement did not approach significance. Clearly, there was less over-all resistance to extinction among *S*s in the present (risk) experiment, but the non-significant interaction indicates that the general shape of the function relating resistance to extinction and percentage of reinforcement was not greatly affected by this over-all difference between risk and non-risk conditions.

TABLE I

COMPARISON OF THE FUNCTION RELATING RESISTANCE TO EXTINCTION AND PERCENTAGE OF REINFORCEMENT IN THE PRESENT EXPERIMENT WITH THE FUNCTION FROM AN EXPERIMENT NOT INVOLVING RISK

Source	*df*	*MS*	*F*
Risk vs. Non-Risk	1	1.358	20.58*
Percentage of Reinforcement	4	.183	2.77†
Risk vs. Non-Risk × Percentage	4	.051	
Within-groups	390	.066	

* Significant at the 1% level. † Significant at the 5% level

(2) Expectancies. The expectancies (verbal statements of confidence of winning on the next play) were Vincentized into thirds of the nine plays of acquisition and into tenths for the plays in extinction for each *S*. The mean expectancies during acquisition and extinction, with percentage of reinforcement as the parameter, are plotted in Fig. 2. The corresponding graph with amount of reinforcement as the parameter is not presented since this variable had little effect on expectancy-scores Descriptively, the

curves of Fig. 2 are very similar to those found in the earlier comparison-study.[5]

To determine the effect of the two main variables on changes in expectancies, two analyses of variance for trend were performed, one over the three blocks of trials in acquisition, the other over the third (last) block of trials in acquisition and the first, fifth, and tenth blocks of trials

TABLE II

ANALYSIS OF TREND OF EXPECTANCIES OVER TRIALS IN ACQUISITION AND EXTINCTION

Source	Acquisition			Extinction		
	df	*MS*	*F*	*df*	*MS*	*F*
Percentage (P)	4	41.28	15.07†	4	30.64	10.60†
Amount (A)	1	.74		1	.78	
P×A	4	10.65	3.89†	4	9.94	3.44†
Between *S*s (error)	290	2.74		290	2.89	
Trials (T)	2	2.73	5.46†	3	86.30	115.07†
T×P	8	4.86	9.72†	12	6.37	8.49†
T×A	2	.59	1.18	3	2.67	3.56*
T×P×A	8	.44		12	.63	
Pooled *S*s×T (error)	580	.50		843‡	.75	

* Significant at the 5% level.
† Significant at the 1% level.
‡ Reduced by 27 *df* (9 *S*s×3 extinction-trials) because 9 *S*s did not complete acquisition, so had no expectancy scores during extinction.

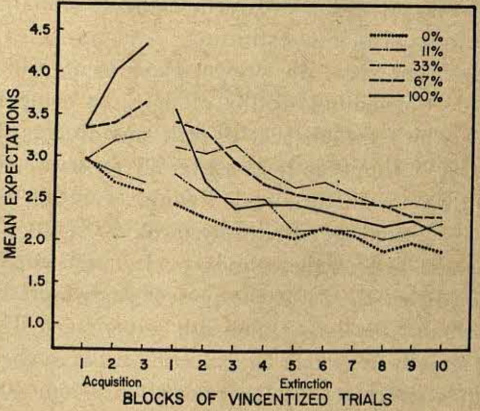


FIG. 2. MEAN EXPECTANCY-SCORES OVER VINCENTIZED BLOCKS OF TRIALS DURING ACQUISITION AND EXTINCTION WITH PERCENTAGE OF REINFORCEMENT AS THE PARAMETER

in extinction. Both analyses are summarized in Table II. The results of both analyses are much the same. As a main effect, amount of reinforcement was not significant. Percentage of reinforcement was significant, as was the interaction of percentage and amount. This interaction was the result of a greater change in expectancy-scores with 1¢ reward than with 5¢ reward over the range from 0% to 100% reinforcement.

---

[5] Lewis and Duncan, *op. cit.*, 1957, 117. The results shown in Fig. 3 include expectancies from groups given all four amounts of reward (1¢, 10¢, 25¢, 50¢), but it was shown that expectancies did not vary significantly as a function of this variable.

The terms of major interest in Table II are those for trials, and in particular the interactions of the main variables with trials. It can be seen that the term for trials and the term for interaction of percentage of reinforcement with trials are highly significant in both analyses. Expectancies changed over the course of both acquisition- and extinction-trials, and changed differentially as a function of percentage of reinforcement. These effects appear clearly in Fig. 2. Expectancies increased during acquisition in the 33%, 67%, and 100% Groups, but decreased in the 0% and 11% Groups. From the end of acquisition to the end of extinction, expectancies decreased in a manner that was almost directly related to percentage of reinforcement; the decrease in mean expectancy was 0.55, 0.53, 0.72, 0.90, and 1.23 for the 0%, 11%, 33%, 67% and 100% Groups, respectively.

In Table II, the interaction of amount of reinforcement with trials was not significant during acquisition, and was significant at only the 5% level during extinction; again, amount of reinforcement had little effect. The triple interaction, among the two main variables and trials, was not significant either in acquisition or extinction.[6]

Discussion. In our view, the present experiment, which involved risk-taking, yielded two findings worthy of note. As compared to an earlier experiment in which the element of risk was absent, there was, in the present study, (a) significantly less over-all resistance to extinction, but (b) no major change in the usual declining function relating resistance to extinction and percentage of reinforcement.

Concerning the first of these findings, less resistance to extinction, it should be noted that the *S*s of the present study were run at a different time than the *S*s of the earlier, non-risk, experiment. Thus, it is possible that the difference in resistance to extinction between the two experiments resulted from a difference in the populations from which the *S*s were drawn. This does not seem likely. The *S*s for all experiments are always drawn from the same introductory courses in psychology, and from all quarters of the academic year. Resistance to extinction in the present experiment was lower than it was in any of the three previous experiments in which the same apparatus was used to study various conditions combined with partial reinforcement, but in which the element of risk was lacking.[7] It seems reasonable to assume that *S*s in the present study were influenced to stop playing the apparatus sooner than usual by the fact that continued playing meant risking the loss of all of their initial stake.

---

[6] These results on expectancy are similar to results in the earlier studies. *Cf.* Table 1, Lewis and Duncan, *op. cit.,* 1957, 118, and Table 2, 119.

All five groups differentiated in terms of percentage of reinforcement showed numerically less resistance to extinction than the groups with which they were compared (see Fig. 1). Despite this, the usual inverse relationship between percentage of reinforcement and resistance to extinction was maintained. Although the relationship (lower curve in Fig. 1) is somewhat irregular, and seems to be slightly flatter than the curves reported in previous studies, there was no clear evidence that the function was much changed by the element of risk. In all situations that we have used with the present or with a similar apparatus, the general relationship between percentage of reinforcement and resistance to extinction has been a very stable phenomenon.

Finally, the results of the present experiment with regard to amount of reinforcement as a variable, and with regard to expectancy as a response-measure, are essentially the same as have been found earlier. Amount of reinforcement has had little influence on resistance to extinction when disks are valued at less than 50¢ each.[8] Expectancies have always changed differentially as a function of percentage of reinforcement during both acquisition and extinction.[9]

SUMMARY

Resistance to extinction in a situation involving some risk-taking was measured as a function of percentage of reinforcement and of amount of reward. Each of 300 *S*s was provided with a limited stake of disks with which to play an electrical slot machine. If on any play the machine did not pay off, *S* lost a disk; if the machine paid off, *S* won an additional disk. For half the *S*s, each disk was worth 1¢, for the other half, 5¢. These two amounts of reward were combined in a factorial design with five percentages of reinforcement; the machine paid off on either 0%, 11%, 33%, 67%, or 100% of the first nine trials. After these, the acquisitional trials, the machine never paid off again (extinction). The *S*s stopped

---

[7] Lewis and Duncan, *op. cit.,* 1957; *J. exp. Psychol.,* 1958; *J. abnorm. soc. Psychol.,* 1958.

[8] Lewis and Duncan, *op. cit.,* 1957, 118.

[9] Lewis and Duncan, *op. cit.,* 1957; *J. exp. Psychol.,* 1958. But note that expectancies may not change if the *S*s have only vicarious experience of winning during acquisition; see Lewis and Duncan, *J. abnorm. soc. Psychol.,* 57, 1958, 324.

playing whenever they wished, at which time they were paid with money for any disks remaining. The response-measures were total number of plays, and *S*'s expectancies, stated in numerical terms, of winning or not winning on each play.

The results were that resistance to extinction, as measured by number of plays, decreased as percentage of reinforcement increased. Resistance to extinction did not vary as a function of amount of reward. Expectancies changed differentially as a function of percentage of reinforcement during both acquisition and extinction. Resistance to extinction was significantly less than had been found in an earlier, comparable study in which the element of risk was lacking, but there was no evidence that risk effected any significant change in the usual inverse relationship between resistance to extinction and percentage of reinforcement.

PAIRED-ASSOCIATE LEARNING AS A FUNCTION OF SIMILARITY: SEMANTIC SIMILARITY BETWEEN STIMULUS- AND RESPONSE-ITEMS

By TAKAO UMEMOTO, Kyoto University, Japan

A number of previous studies have been concerned with semantic similarity in paired-associate learning.[1] Their findings have consistently supported the hypothesis of Gibson that, due to interferences through generalization, the more similar items there are in a list, the more difficult it is to learn that list.[2] They find also that similarity among stimulus-items is more influential than similarity among response-items. Most of the experiments have been limited to varying similarity among stimulus-items or among response-items. One of the exceptions is the experiment of Underwood in which he constructed two lists of paired-associates, with two degrees of similarity.[3] In the high-similarity list, five sets of four highly synonymous adjectives were used as both stimuli and responses. He found that the high-similarity lists took many more trials to learn than did the low-similarity lists. Thus, to the previously studied variables of similarity among stimuli and among responses is added a variable of stimulus-response similarity, a variable which may be designated inter-S–R similarity.

According to Gibson's hypothesis, increasing the similarity among stimulus-items increases the difficulty of mastery because of the confusion that occurs through stimulus-generalization, one stimulus erroneously substituting for another; this hypothesis has been generally supported. While the hypothesis in its original form ended here, it may be conjectured, by analogy, that if response-items are similar, they, too, might be confused through generalization; a corresponding confusion might occur if some stimulus-items are very similar to some response-items. Underwood found, when both stimulus- and response-lists consisted of heterogeneous adjectives with some stimulus-items highly similar to some response-items, that interference did indeed occur and that mastery was more difficult than when this inter-S–R similarity was lacking. The present experiments move on from this design

---

* Received for publication July 28, 1960.

[1] B. J. Underwood, Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity, *J. exp. Psychol.,* 42, 1951, 153-161; Studies of distributed practice: VIII. Learning and retention of paired-adjectives as a function of intra-list similarity, *ibid.,* 45, 1953, 143-149; Intra-list similarity in verbal learning and retention, *Psychol. Rev.,* 61, 1954, 150-166; Y. Morikawa, Studies in paired-associates learning: IV. On the influence of intra-stimulus and intra-response similarity upon learning and recall, *Jap. J. Psychol. Rev.,* 3, 1959, 116-127.

[2] E. J. Gibson, A systematic application of the concepts of generalization and differentiation to verbal learning, *Psychol. Rev.,* 47, 1940, 196-229.

[3] Underwood, *op. cit.,* 1951, 153-161.

to introduce additional aspects of similarity, found when *both* stimulus- and response-items are selected from a common pool of similar items, so that inter-S–R similarity is added to intra-S and intra-R similarity.

There are several methods of controlling inter-S–R similarity. In a previous investigation, we varied the number of items in common on the stimulus- and response-sides of paired items.[4] Other methods include variation of formal similarity (*e.g.* the number of common letters in the items), and semantic similarity, that is, similarity in meaning and familiarity of items. The present paper is concerned with inter-S–R similarity in which there is also intra-stimulus and intra-response semantic similarity. Two experiments are reported, Experiment I conducted on Japanese students in Japan, Experiment II, on American students in the United States.

EXPERIMENT I

There are three conditions in which inter-S–R similarity was varied from high, to medium, to low. The lists to be learned were presented by means of a Takei-type memory-drum at a rate of 2 sec. for each exposure, with 2 sec. between items and between trials. To avoid serial effects, the pairs were presented in four random orders.

Method. The method of anticipation was used, in which *S* had an opportunity to respond to the stimulus-word before the response-word appeared. Learning trials were continued until *S* reached a criterion of one perfect trial. Before the experiment began, *S* was given ordinary instructions about the anticipation-method, but he was not informed about the purpose of the experiment or the similarity of lists.

Subjects. Nine students in a psychology course at Kyoto University served as *Ss*, each *S* serving in all conditions according to a 3 × 3 Latin-square design. They had completed an introductory course in experimental psychology, and had had some experience in experiments in memorization. An interval of at least a day was allowed between each condition for each *S*.

Material. One type of list was used for each condition, but there were three lists of each type. Each list contained 8 pairs of 16 Japanese adjectives taken from a table prepared by Umemoto, Morikawa, and Ibuki.[5] In this table, 1190 adjectives and their associative frequencies are recorded. The associated adjectives were obtained as response-words under instructions to give similar associations to 120 basic Japanese adjectives. The similarity between the stimulus- and response-adjectives was later rated by university students on a 10-point scale; these ratings are listed, along with a corresponding rating of familiarity. In this experiment, the list for the high-similarity condition was composed entirely of adjectives which appeared as associated responses to the same key-word. In the medium-similarity lists, the adjectives on the stimulus-side all belonged to one group, that is, they were responses given as associates to the same key-word, and the adjectives on the response-side all belonged to one group, but as responses to another key-word. The two key-words defining the groups of responses were, however, also somewhat similar, in that at least one of these two appeared as an associated response to the other as a key-word, though at low frequencies in the table of frequency of association. For instance, to the key-word 'atatakai' (warm) appeared the associated response-words 'ondanna,' 'shinsetsuna,' 'nukui,' etc., while to the key-word 'akarui' (bright) occurred the associated adjectives 'hogarakana,' 'mabushii,' 'akirakana,' etc.; but among the associated responses to 'atatakai' (warm) is also found the adjective 'akarui' (bright), though the rated similarity on the 10-point scale is but 1.0. Finally, for the low-similarity lists, the adjectives on the stimulus-side all belonged to one group, and those on the response-side to another group, but neither of the two key-words was ever given as a response to the other.

---

[4] Takao Umemoto and E. R. Hilgard, Paired-associate learning as a function of similarity: Common stimulus and response items within the list, *J. exp. Psychol.*, 62. 1961, 97-104.

[5] Takao Umemoto, Y. Morikawa, and M. Ibuki, Similarity and familiarity scales of Japanese adjectives, *Kyoto Univ. Stud. Educ.*, 1, 1955, 85-116.

TABLE I

MEAN SCORES OF SIMILARITY AND FAMILIARITY FOR LISTS OF EXPERIMENT I
(Ratings on a 10-point scale)

Type of List (Inter-S–R similarity)	Intra-S similarity	Intra-R similarity	Familiarity of S	Familiarity of R
List 1 (High)	6.2	6.1	7.7	7.0
List 2 (Medium)	6.4	6.4	7.7	7.5
List 3 (Low)	6.9	5.9	7.6	7.5

The success with which the lists were equated for similarity of the adjectives composing each list, and for the familiarity of these adjectives, is shown by the mean values of Table I. There were no significant differences among the ratings of either the similarity or the familiarity of the stimulus- and response-words making up the paired-associate lists.

Results. The mean number of trials required to reach a criterion of one perfect trial for the conditions of high, medium, and low similarity was 22.8, 25.7, and 36.9, respectively. An analysis of variance based on these scores revealed that only the differences among types of list were significant at the 1% level ($F = 12.35$, with 2 *df.*). Furthermore, significant t-scores were obtained for the differences between conditions of high and low similarity ($t = 3.30$), and medium and low similarity ($t = 3.26$), though the difference between high and medium similarity was not significant ($t = 1.42$). The mean number of overt errors was 31.3, 39.2 and 57.1 for high, medium, and low similarity, corresponding to the results for trials, though the percentages of overt error to total errors including silent responses was essentially the same (35.0%, 34.4% and 32.2%) for high, medium, and low similarity respectively. As indicated in Table II, the main source of the overt errors was almost always within response-items; intrusions from the stimulus side into response were very few. There was a slight tendency to increase the intrusion-error from stimulus to response as inter-S–R similarity increased, though the tendency was not significant. Also, a measure of oscillation (range between first success and the last failure) showed similar results, the mean for List 1 being 60.1; for List 2, 71.8; and for List 3, 119.4. All these error-scores (overt errors, failures to respond, and duration of oscillation) showed the same results as the trials to mastery: the more similar the stimulus and response members of a pair are to each other, the more rapid the learning and the fewer the errors. The results will be discussed after reporting Experiment II.

TABLE II

SOURCE AND TYPE OF OVERT ERRORS FOR LISTS OF EXPERIMENT I
(Mean cumulative errors per list)

Type of List (Inter-S–R similarity)	Intrusions from R	Intrusions from S	Extra-list intrusions	Fractional responses
List 1 (High)	23.9	2.0	0.8	4.5
List 2 (Medium)	32.7	0.7	0.7	5.2
List 3 (Low)	48.1	0.9	0.0	8.1

EXPERIMENT II

Sometimes attempts to replicate an experiment in a different country with different materials and *S*s have failed, even when the experiments were similarly designed. It was therefore decided to do a similar experiment in the United States with American students, to determine how generalizable the results are across cultural and linguistic differences.[6] Some differences in conditions, beyond differences between Japan and America, may be noted. The American *S*s, from an introductory course in psychology, had not had experience with memorization in the laboratory, while the Japanese *S*s were experienced. The Japanese *S*s participated in all three conditions; each American *S* participated in only one of the conditions. The availability of adjectival materials led to a somewhat different method of constructing lists. That these changes were important but not crucial will be evident when the results are presented.

Method. As in Experiment I, there were three conditions: the first with high inter-S–R similarity, the second with medium similarity, the third with low similarity. After learning a practice list of three pairs to a criterion of one perfect, each *S* then learned one list of paired-associates in another experiment, and then went on to learn the list in this experiment. Trials were continued to one perfect trial. The experiment was so planned that there would be as few interferences as possible between the conditions of this and the prior experiment. Usual instructions about the anticipation-method were given before the experiment began. After reaching the criterion of one perfect trial, *S* was tested on backward association by being asked to give the stimulus-word associated with a response-word, each response-word being written on a card and shown to *S*.

---

[6] Experiment II was conducted at Stanford University during 1959-1960, while the author was there as a visiting scholar.

Subjects. The numbers of *S*s completing the experiment were, for the three groups, 11, 10, and 11 respectively.[7] They were all university students, and as stated above, they served as *S*s for another experiment before the present one.[8]

Materials. The first two lists used the same stimulus-words that were chosen by Underwood for his study of intra-list similarity.[9] He had one series of adjectives, all of which had unpleasant meanings. This group of adjectives was divided into two, one half to be used as the stimulus-words, one half of the response-words in List 1, that of high S–R similarity. The same response-words were used for all lists (List 1, 2, and 3) because previous studies have shown that changing the response-words has a greater effect upon paired-associate learning than changing the stimulus-words,[10] and this was not the variable important in this experiment. The stimulus-adjectives in List 2, also from Underwood's experiment, were all adjectives having a pleasant meaning. Thus the S–R relationship in List 1 was unpleasant-unpleasant, while in List 2 it was pleasant-unpleasant. In List 3 the stimulus-adjectives were all similar in having the meaning of 'width.' Thus the relationship between S–R pairs for List 3 was width-unpleasant, making List 3 semantically the least similar of the three lists. In arranging the pairs of words care was taken to avoid any apparent relation, such as common meanings or common parts, *e.g.* hurtful-harmful or hurting-shocking. While an American graduate student assured the writer that the words were all very familiar to college students, a later check against the Thorndike-Lorge word-list showed the words referring to width to be somewhat more frequent in usage than the selected adjectives.[11] The lists used are given in Table III.

TABLE III

LISTS USED IN EXPERIMENT II

List 1 High inter-S–R similarity	List 2 Medium inter-S–R similarity	List 3 Low inter-S–R similarity
wicked-painful	elated-painful	wide-painful
dreadful-nasty	carefree-nasty	broad-nasty
ugly-shocking	gleeful-shocking	extensive-horrid
filthy-tragic	blissful-tragic	spacious-tragic
depraved-harmful	festive-harmful	roomy-fearsome
hateful-fearsome	happy-fearsome	large-harmful
hurting-evil	hearty-evil	huge-evil
cruel-horrid	jolly-horrid	vast-shocking

---

[7] Technical difficulties made data from 5 *S*s unusable, resulting in the unequal numbers in the experimental groups.

[8] Umemoto and Hilgard, *op. cit.*, 1961, 97-104.

[9] Underwood, *op. cit.*, 1951, 153-161.

[10] Umemoto, On the relative weight of stimulus versus response words in rote learning, *Jap. J. Psychol.*, 21, 1951, 46-55; Morikawa, Studies in paired-associates learning: I. Forward-backward recall gradient, *Jap. J. Psychol.*, 26, 1955, 156-171; G. Mandler and E. H. Campbell, Effect of variation in associative frequency of stimulus and response members on paired-associate learning, *J. exp. Psychol.*, 54, 1957, 269-273.

[11] E. L. Thorndike and Irving Lorge, *The Teacher's Word Book of 30,000 Words*, 1944.

Results. Performance on the practice-list showed the three groups to be well equated, the mean trials to mastery being 4.00, 4.40, and 4.36 for the three groups learning Lists 1, 2, and 3, respectively. The F-score was 0.79 ($df. = 2$ and 29), and therefore not significant.

The mean number of trials to reach a criterion of one perfect trial for Lists 1, 2, and 3 was 14.9, 24.5, and 18.8, respectively. Again List 1, with highly similar items is the most rapidly mastered, but now List 2, with medium similarity, is the most difficult. Thus, in relation to Experiment I, Lists 2 and 3 have changed places. An analysis of variance, based on the square roots of the numbers of trials to mastery, yielded an F-score of 3.79 ($df. = 2$ and 29) which is significant for Lists at the 5% level. The t-tests showed, however, that the only significant difference was between the scores for Lists 1 and 2 ($t = 2.57, p < 0.05$).

A study of the cumulative errors to successive criteria yielded essentially the same result. The analysis of variance for the error-scores revealed that only the variable of lists was significant; Stages of learning and the interaction of these variables were not significant. Again List 1 was the easiest, List 2 the hardest, with List 3 falling between them. The largest proportions of overt errors were again intra-response intrusions (14.9, 23.8, and 15.8 for List 1, 2, and 3), and few errors had their source on the stimulus-side (0.54, 1.09, and 0.54 for Lists 1, 2, and 3).

We have found in a previous study that backward recall, *i.e.* R–S recall, following a criterion of one perfect trial, varied inversely with difficulty in learning.[12] In this experiment, however, the mean correct backward associations for the three lists, which were 6.7, 7.2, and 7.1 for Lists 1, 2, and 3, were not significantly different from each other.

DISCUSSION

The results of Experiment I, performed in Japan, and of Experiment II, performed in America, agree in showing that with high inter-S–R similarity paired-associate lists are easier to learn than those with lower degrees of inter-S–R similarity. This finding contradicts the hypothesis originally stated. The disagreement between Experiment I and Experiment II, which resulted in a reversal of difficulty between Lists 2 and 3, can be explained in part on the basis of the greater familiarity of the 'width' adjectives in List 3 in Experiment II.

How is the finding to be explained that the most similar pairs were the most easily acquired? It is obvious that Gibson's theory, in its original form, cannot be applied to this experimental situation, since her theory is stated only in terms of stimulus-generalization.[13] If we extend her hypothesis, however, to the present situation of stimulus-response similarity, we could expect more interferences for the list having more inter-S–R similarity because the similarity disturbs differentiation of stimuli and responses in the list. The probability of such disturbance will increase as the inter-S–R similarity increases. Our results did not confirm this conjecture. As reported in the results of Experiments I and II, a large part of overt errors were intra-response intrusions, and few errors were from stimulus-side. Furthermore those few intrusions from stimulus-items did not always increase as the inter-S–R similarity increased. From these results we can say that by contrast with our previous study, in which an increase of identical stimulus- and response-items had an interfering effect upon paired-associate learning, increase of inter-S–R semantic similarity does not interfere with the learning process.

---

[12] Umemoto and Hilgard, *op. cit.*, 1961, 97-104.

Köhler once tried to extend his theory of perceptual organization to paired-associate learning, showing that similar (homogeneous) pairs were much more readily recalled than dissimilar (heterogeneous) ones.[14] Postman and Riley attacked Köhler's theory and experiments on several points, and showed the possibility of an explanation according to associationistic principles of learning.[15] In their paper they proposed an hypothesis of 'initial bias' in response-selection, stating that "*S*'s previous experiences with sequences of syllables, numbers, etc., will result in a significant bias for a rule of response selection, viz., the rule that like goes with like."[16] Though the Postman-Riley hypothesis of 'initial response bias' was concerned only with similarity of materials, such as number, syllable or figure, which are placed into a single list, we might expect a similar kind of bias for semantic similarity. Then the present findings can be interpreted as showing that such bias is capitalized upon in learning, and can facilitate the learning of a list having similar items on the stimulus- and response-sides. Mandler has stressed the importance of "differentiating responses" in determining the transfer effects within a list.[17] Thus all the stimulus-words may be differentially categorized as 'pleasant,' 'unpleasant,' or

---

[13] Gibson, *op. cit.*, 196-229.

[14] Wolfgang Köhler, On the nature of associations, *Proc. Amer. Philos. Soc.*, 84, 1941, 489-502.

[15] Leo Postman and D. A. Riley, A critique of Köhler's theory of association, *Psychol. Rev.*, 64, 1957, 61-72.

[16] Postman and Riley, *op. cit.*, 66.

[17] George Mandler, Response factors in human learning, *Psychol. Rev.*, 61, 1954, 235-244.

'width' words. If now the differentiated response 'unpleasant' applies to both stimulus- and response-words (as it would in Experiment II for the condition of highest inter-S–R similarity) this would narrow the possible responses to those conforming to the initial response-bias toward like-leads-to-like. Thus the high S–R similarity-lists might be the more readily learned, the factor of response-bias overriding any interference due to generalization.

The differences between previous findings, those of Underwood,[18] and our own,[19] and the present findings call for an explanation. While further experimentation is needed to explain the differences in detail, there are significant ways in which the experiments differ from each other.

(1) Underwood's experiment differed from ours in that all pairs were heterogeneous, even in the condition of high inter-S–R similarity. Thus one pair was 'dirty-frigid,' while another was 'arctic-devoid.' Thus the similarity between 'frigid' and 'arctic' could not provide any advantage in learning through response-bias toward similarity. In our high-similarity conditions the similarity existed also between the paired items.

(2) Our earlier experiments differed from the present ones in that the inter-S–R similarity was controlled through the number of identical items, while the present experiments used semantic similarity. Apparently the confusion between *identical* stimulus- and response-items produces far more interference than that between *semantically similar* stimulus and response items.

It is apparent that specific differences in the arrangement of items and in the definition of similarity need to be recognized before general statements can be made about the place of similarity in paired-associate learning.

SUMMARY

Two experiments are reported, one performed in Japan, and the other in the United States. The intra-stimulus and intra-response similarities of lists of paired-associate adjectives were kept constant, but the similarity between stimulus- and response-items was varied from high similarity through medium similarity to low similarity. In the first experiment the degrees of similarity were controlled by using a standardized association-table. In the second experiment the first group of items all had unpleasant meanings on both stimulus- and response-sides, the second group combined pleasant stimulus-items with the same unpleasant response-items,

---

[18] Underwood, *op. cit.,* 1951, 153-161.
[19] Umemoto and Hilgard, *op. cit.,* 1961, 97-104.

while the third group used adjectives referring to width combined with the unpleasant response-items.

In both experiments the lists with the most similar S–R items were the easiest to master. The reversal in the relative positions of the other two groups is attributed in part to a faulty construction of the third list in the second experiment. The results are interpreted on the basis of the Postman-Riley hypothesis of initial response-bias acting upon differential responses to the similar stimuli.

ESTIMATED PROBABILITY OF SUCCESS AS A FUNCTION OF VARIABILITY IN PERFORMANCE

By JAMES C. DIGGORY and BENA OSTROFF, University of Pennsylvania

Using a card-sorting task with experimentally controlled performance-curves and clearly defined goal and deadline, Diggory, Riley, and Blumenfield showed that *S*'s estimate, on any trial, of the likelihood that he will succeed within the remaining trials depends on (a) his rate of improvement, (b) whether or not the rate is changing, (c) his mean level of performance, (d) the number of trials remaining (distance from deadline), and (e) the value of the outcome of a successful performance.[1] Estimates of probability of success, $P(s)$ generally decreased from trial to trial, especially near the deadline, although there were significant differences among the means and slopes of individual $P(s)$-curves, differences which seemed to deserve further study. These differences are dealt with here, but the main purpose of the present paper is to clarify the effect on $P(s)$-estimates of real, or expected variability in performance.

The problem arose in the following way: One of the performance-curves in the previous experiments was of the same form as Part 1 of Fig. 1, leveling off at 39, just one point below the specified goal. Some *Ss* to whom this curve was represented as their own kept their $P(s)$-estimates very high throughout the series, *contrary* to the general trend of lowering them as the deadline was neared. One plausible explanation of this behavior is that these *Ss* expected some irregularity in their performance curves—an upward deviation of only one unit would spell success. In fact, some *Ss* said they regarded their stated performance-curve as a trend line around which they expected some variation to occur. If this is the correct explanation, then *Ss* in the same conditions who did not keep their $P(s)$-estimates high either did not expect fluctuations in their performance-curves or expected they would not be favorable ones. The chief manipulation in the present experiment is designed to control the expectation of variability, at least in the later trials, building irregularity into the rigged performance-curves. Our hypothesis is that variability or irregularity in the history of generally improving performance would produce higher $P(s)$-estimates, or more retardation in their decline near the deadline than would a highly predictable, regular performance-history, and this should be especially true when *S*'s performance is near the goal.

If *S* views his performance-curve as a 'trend line,' then a highly regular

---

* Received for publication September 13, 1960. This work was supported by a grant from the University of Pennsylvania Committee on the Advancement of Research.

[1] J. C. Diggory, E. J. Riley, and Ruth Blumenfeld, Estimated probability of success for a fixed goal, this JOURNAL, 73, 1960, 41-55.

curve allows him a more accurate prediction of its future course than would an irregular curve. This is a form of the 'expanded judgment' situation studied by Irwin and his students.[2] They showed that a naïve *S*'s confidence in his judgment that two means differ, or that one mean differs from zero, varies inversely with the *SD* of the presented data. Applying these results to our situation leads to the conclusion that the more smooth and regular the progress of performance, the more confidently *S* will predict his own failure; but, on empirical grounds, we expect *S*s with fluctuating performance-curves, in spite of the relative difficulty of predicting the course of their improvement, to make a constant error in the direction of anticipating success rather than failure. In this connection, we should point out that Festinger, discussing judgments in social-comparison processes, found it necessary to explain differences between judgments of ability and judgments of opinion by invoking the notion of "an upward push" with respect to abilities.[3]

Presenting irregular performance-curves to the *S*s also gave the opportunity to examine whether a single large increment in performance would affect $P(s)$ differently if it occurred far from the deadline rather than near it. This manipulation is described below under 'method.'

METHODS

Eighty undergraduate university men were randomly assigned, 4 each, to the 20 experimental conditions determined by the characteristics of the performance-curves in Fig. 1. These curves were controlled, as in our previous experiments, by deceiving *S* about the time for each trial, stopping him when he had sorted a predetermined number of cards.

Variability or irregularity of performance-curves was introduced by drawing envelopes of ±1 and ±2 units amplitude around each of the smooth curves shown in Parts 1-4 of Fig. 1. From within each envelope, points were selected to form the irregular curves, subject to the restrictions that they (a) were generally rising, (b) had as few reversals as possible, and (c) had the same means as the corresponding smooth curves. Thus, there were three levels of irregularity in this sense: zero, ±1, and ±2.

Position of the 'maximal jump,' early or late in the series of trials, was determined for each of the irregular curves by locating the maximal increment (from the lower to the upper limit of the envelope) either between Trials 3 and 4 or between Trials 6 and 7. This is not necessarily the largest increment in a performance-curve; it simply marks a point where there is a sharp upward deviation

---

[2] F. W. Irwin, W. A. S. Smith, and J. F. Mayfield, Tests of two theories of decision in an 'expanded judgment' situation, *J. exp. Psychol.*, 51, 1956, 261-258; Irwin and Smith, Further tests of theories of decision in an 'expanded judgment' situation, *ibid.*, 52, 1956, 345-348; Value, cost, and information as determiners of decision, *ibid.*, 54, 1957, 229-232.

[3] Leon Festinger, A theory of social comparison processes, *Hum. Relat.*, 7, 1954, 117-140.

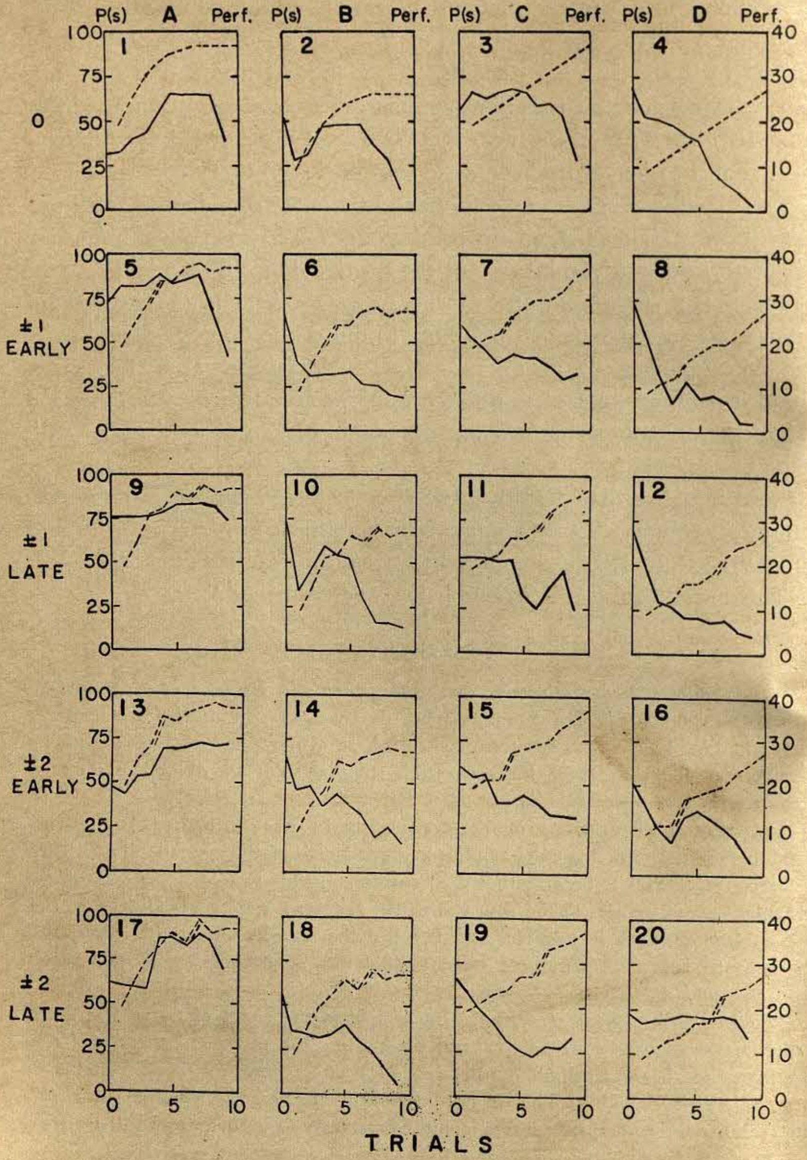


FIG. 1. PERFORMANCE-CURVES AND CORRESPONDING MEAN ESTIMATED *P(s)*-CURVES FOR THE EXPERIMENTAL CONDITIONS

(The dashed lines are performance-curves, to be read on the right-hand scale of each graph. The solid lines show the means of estimated *P(s)* for four *Ss* in each condition, to be read against the scale at the left of each graph. The numbers 0, ± 1, and ± 2 at the left of the figure indicate the amount of variability in the performance-curves; 'Early' and 'Late' refer to the position of the 'maximal jump' which is shown also by the doubled line in each dashed curve.)

in the general performance-trend. The position of the maximal jump is shown by the double-lined section of each performance-curve in Fig. 1.

Performance-level was controlled by generating two sets of curves, the first beginning at 19 points and ending at 37 points (Sets A and C of Fig. 1), the other set being exactly 10 points lower every trial (Sets B and D, Fig. 1).

The exact instructions to the *S*s, a photograph of the display 'graph,' and the $P(s)$ rating-scale have been presented earlier.[4]

RESULTS

The solid lines of Fig. 1 locate the mean $P(s)$-estimates made by *S*s in the various conditions. Though the initial (Trial 0) estimates are quite diverse, analysis of variance reveals no significant differences among them either in main effects or interactions. Analysis of the data over all trials by Alexander's trend-test shows that for the entire set of data, the slope of the best fitting straight line is significantly negative.[5] Further, the mean $P(s)$ and slope for the high-performance conditions are significantly greater than for low performance, confirming our earlier results. Since there were no significant differences between the ±1 and ±2 variable-performance conditions, they are hereafter treated together as the irregular performance-conditions. Since the 'maximal jump' conditions produced only local reversals in the downward trends of the $P(s)$-curves, without affecting over-all means or slopes, they will not be discussed further.

TABLE I

CHARACTERISTICS OF $P(s)$-CURVES AS A FUNCTION OF ARTIFICIALLY INDUCED CHARACTERISTICS OF PERFORMANCE-CURVES

	Condition				*P*
	A	B	C	D	
Over-all mean performance	32.6	22.6	28.0	18.0	
Mean initial $P(s)$	58.2	64.1	59.1	65.2	>0.05*
Mean $P(s)$/subject/trial	69.1	41.9	45.4	32.9	<0.01†
$P(s)$-slope	−1.48	−3.98	−3.16	−4.76	<0.01†
Mean final $P(s)$	59.7	13.0	30.5	12.2	<0.01*

* Analysis of variance.

† Kruskal-Wallis one-way nonparametric analysis.

Table I summarizes the important characteristics of the $P(s)$-curves as they relate to Conditions *A, B, C,* and *D.* As Fig. 1 shows, Conditions *A* and *B* include the negatively accelerated performance-curves, while those of *C* and *D* have constant slope. The high curves are in Conditions *A* and *C*, the lows in *B* and *D*. From Table I it is evident that mean initial $P(s)$-

---

[4] Diggory, Riley, and Blumenfeld, *op. cit.*, 44-47.

[5] H. W. Alexander, A general test for trend, *Psychol. Bull.*, 43, 1946, 533-557.

estimates are independent of the experimental treatments, as they should be, but the slopes and over-all means of the $P(s)$-curves are significantly related to the experimental conditions. Of course, the over-all means and the slopes are not independent, provided the curves start at approximately the same level and the slopes are approximately linear. They are presented together here to indicate that indeed they are not independent. Since we are interested primarily in the effects of our experimental conditions on the trends of change in $P(s)$, we shall limit the following discussion to that feature of the $P(s)$-curves. The slopes are those for the best-fitting straight lines, which is an over-simplification of most of the $P(s)$-trends, but quite adequate to detect the differences required to test our hypothesis.

TABLE II

SLOPES OF $P(s)$-CURVES, TRIALS 5–9, AND FINAL $P(s)$-ESTIMATES FOR REGULAR AND IRREGULAR PERFORMANCE-CURVES

Condition

Performance	A		B		C		D	
	slope	final mean	slope	final mean	slope	final mean	slope	final mean
Regular	−5.40	40.0	−9.38	10.5	−8.02	29.8	−8.38	18.8
Irregular	−3.86	64.6	−6.53	13.7	−1.08	30.7	−4.11	24.3

Variability. The effects of variability on the $P(s)$-trends are presented in Fig. 2, which shows that introducing irregularity into the performance-curves retards the decrease in $P(s)$-estimates in the later trials. As Table II shows, the final mean estimates are generally higher for the irregular performance-curves than for regular ones. The slopes of the best fitting straight lines are significantly more negative in the last five trials for the regular performance conditions than for the irregular ones ($p < 0.01$). Furthermore, in the regular conditions, every *S* shows a greater negative change in the second half of his $P(s)$-estimates than in the first half, but this is true of only 38 of the 64 *S*s in the combined irregular conditions. These findings confirm our hypothesis that a history of variability in performance will retard the decline in $P(s)$-estimates which generally occurs with approach to a deadline.

Individual differences. The effects of individual differences in initial $P(s)$-estimates are shown in Fig. 3. In Condition *A*, especially, initial mean $P(s)$ for the smooth performance-curve was markedly lower than for the irregular curves. Since this made straightforward interpretation of the results difficult, we decided to see whether the trends in $P(s)$ showed any systematic differences depending on whether *S*'s initial estimate was above or below the median for his experimental condition. Fig. 3 presents the results of this analysis for the 'irregular' conditions. In all four sets of conditions (*A, B, C. D*) the $P(s)$-curves with the higher initial level fall more rapidly than those with the lower initial level.[6] Analysis of comparable data from 'regular' conditions in one of our previous experiments shows the same result.[7] This effect does not come from initial $P(s)$-estimates being so low as to leave no room for further decline, since one of the

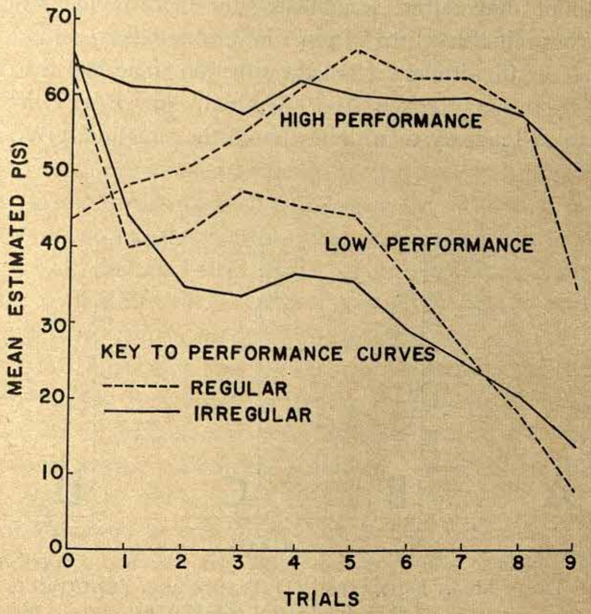


FIG. 2. MEAN ESTIMATED $P(s)$ AS A FUNCTION OF IRREGULARITY AND MEAN LEVEL OF PERFORMANCE

(The curves labelled High Performance are averages of the $P(s)$-curves from Columns *A* and *C* of Fig. 1; Low Performance curves are from Columns *B* and *D* of Fig. 1.)

low-starting curves (Part *A*) has a generally positive slope (+2.31) with plenty of room to fall at an even greater rate than the corresponding high-starting curve (slope = −0.75); the low-starting curve in Part *C* (slope = −1.35) has plenty of room to fall at the same rate as the high-starting curve (slope = −4.02); and in Conditions *B* and *D* (low average performance-level) the high-starting curves end significantly lower than the low-starting curves.

---

[6] $P < 0.01$ in every case, by Alexander's trend-test (*op. cit.*, 533-557).

[7] Diggory, Riley, and Blumenfeld, *op. cit.*, 41-45, esp. 48-49.

The importance of this finding lies in its relation to assertions that people with high self-esteem tend to withdraw more from situations threatening to their self-esteem than do people with low self-esteem.[8] The relevance of our observation to this issue depends upon the assumption that, in the restricted situation of our *S*s, an initially high $P(s)$-estimate corresponds to a high opinion of one's capacity to accomplish the given task while a lower $P(s)$-estimate reflects less confidence. Thus, as the evidence of impending failure accumulates, the more rapid decline of $P(s)$ from a higher initial level than from a lower one reflects a more rapid gain in *S*'s tendency to withdraw from the situation if he had the opportunity to do so. By the same token, *S*s with initially low $P(s)$-estimates would show a weaker tendency to withdraw from the situation, given the chance to do so.

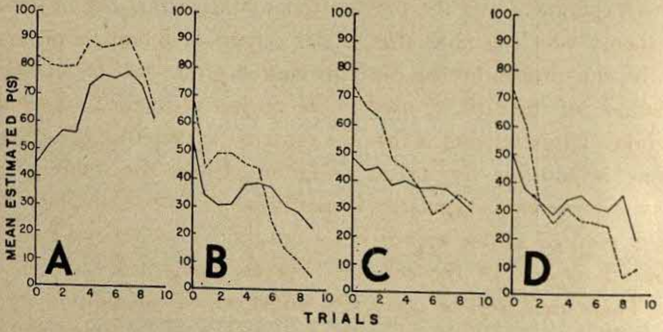


FIG. 3. MEAN ESTIMATED $P(s)$ TRENDS AS A FUNCTION OF LEVEL OF INITIAL ESTIMATE

(These curves refer only to the irregular performance-conditions. The dashed lines connect the mean $P(s)$-estimates of *S*s whose initial estimate is above the median of their group. Solid lines show the trends for those who start below the median.)

---

[8] Saul Rosenzweig, An experimental study of repression with special reference to need-persistive and ego-defensive reactions to frustration, *J. exp. Psychol.*, 32, 1943, 64-74; Saul Rosenzweig and Gwendolyn Mason, An experimental study of memory in relation to the theory of repression, *Brit. J. Psychol.*, 24, 1934, 247-265; T. G. Alper, Memory for completed and incompleted tasks as a function of personality: an analysis of group data, *J. abnorm. soc. Psychol.*, 41, 1946, 403-420; The interrupted task method in studies of selective recall: A reevaluation of some recent experiments, *Psychol. Rev.*, 59, 1952, 71-88; C. W. Erikson, Defense against ego-threat in memory and perception, *J. abnorm. soc. Psychol.*, 47, 1952, 230-235; A. R. Cohen, Experimental effects of ego-defense preferences in interpersonal relations, *ibid.*, 52, 1956, 19-27.

SUMMARY AND CONCLUSIONS

The results of this experiment support the findings of an earlier study that, for given performance-curves, estimates of the probability of success decline with increasing nearness to the deadline. In addition, the present experiment shows that this rate of decrease is smaller when performance approaches the goal at an irregular rate. Early or late occurrence of a large isolated increment in performance produced only local 'tracking' of the performance-curve by the *P*(*s*)-estimates; no other characteristics of the *P*(*s*)-trends are affected by it. The latter finding indicates that estimated *P*(*s*) probably is based on the general aspects of performance: nearness to the goal, nearness to the deadline, and shape of the performance-curve, including its irregularity. We also obtained evidence that series of *P*(*s*)-estimates which begin higher than the median of the general group descend more rapidly and end lower than those which start at a lower level. The relation of this to the alleged tendency for persons with high self-esteem to defend against threats to their self-esteem by withdrawal is discussed.

This experiment and the previous ones were carried out in the context of a theory which assumes that, under certain conditions, a person's estimate of the probability of his achieving a given goal is related to his evaluation of the capacity used in the pursuit of the goal. An important condition of this relation is that the capacity be exercised where its power is alone responsible for success or failure. Under the conditions of the present experiment, irregularity in performance reduces the rate at which *P*(*s*) is reduced as one approaches a deadline; however, we have not explored all aspects of the relation between performance variability and *P*(*s*). There may well be conditions under which variability is an indication of inadequate capacity and *P*(*s*) declines more rapidly than without variability.

SERIAL POSITION AS A CUE IN LEARNING

By SLATER E. NEWMAN, North Carolina State College and
ELI SALTZ, Wayne State University

Research in verbal learning has demonstrated that variables which affect performance in serial learning may not have the same effect on paired-associate learning. Thus, Underwood has reported a direct relationship between intertrial interval and rate of serial learning, but no relationship between intertrial interval and rate of paired-associate learning.[1] Newman and Saltz, have also shown that the Von Restorff effect appears to be more pronounced when the isolated term serves as a response-term in a serial list, but as a stimulus-term in a paired-associate list.[2]

It seems reasonable to assume that these differences in findings derive from differences in the characteristics of the two tasks. Among the differences between tasks which may be of importance is the order in which the items appear during training-trials. In serial learning each item appears in the same serial position on each trial, whereas in paired-associate learning, each pair of items appears in several different orders during the course of learning. Since serial position is more likely to serve as a cue in serial learning than in paired-associate learning, the present experiment was undertaken to study the role of serial position as a cue.

This problem has not been intensively investigated. An approach to the problem, however, was suggested by a procedure employed by McGeoch and McKinney who compared the performance of two groups: a Serial Group for which on all five pairing trials, each of the S–R pairs appeared in the same serial position, and a Random Group for which on all five pairing-trials, each S–R pair appeared in a differential serial position.[3] During each pairing-trial, only the S–R pairs appeared. On the retention-test (the sixth trial) each of the S-terms was again presented in the same serial order for the Serial Group, whereas for the Random Group the S-terms were presented in an order different from any of those during the pairing-trials. Though performance of the Serial Group in terms of number of correct responses was higher than for the Random Group, it was not possible to determine whether this occurred (a) as a function of serial presentation during the pairing-trials,

---

* Received for publication October 25, 1960.

[1] B. J. Underwood, Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intralist similarity, *J. exp. Psychol.*, 44, 1952, 80-87; Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity, *ibid.*, 45, 1953, 133-142.

[2] S. E. Newman and Eli Saltz, Isolation effects: Stimulus and response generalization as explanatory concepts, *J. exp. Psychol.*, 55, 1958, 467-472.

[3] J. A. McGeoch and Fred McKinney, Studies in retroactive inhibition, VIII: The influence of the relative order of presentation of original and interpolated paired associates, *J. exp. Psychol.*, 20, 1937, 60-83.

(b) as a function of the fact that the S-terms for the Serial Groups were presented in the same order during the test of retention as during the pairing-trials, or (c) as a joint function of both. A similar difficulty appears in interpretation for the results of a study by McGeoch and Underwood reporting facilitation for serial as contrasted with random presentation of S–R pairs. Their use of the anticipation method (S-term, then S–R pair) does not permit the clear-cut identification of the source of such facilitation.[4]

To deal with this problem in the present study, a third group (the SD-Group) was employed in addition to the Serial (hereafter referred to as the SS-Group) and Random Groups used in the study of McGeoch and McKinney. For the SD-Group the pairs appeared in the same order on all five pairing-trials, but in a different order on the testing-trial. Addition of this group made it possible to test the prediction that the number of correct responses by the SS-Group on the test of retention (1) would exceed that for the SD-Group, and (2) would exceed that for the Random Group. No prediction was made comparing the SD- and the Random-Groups since it did not seem possible to determine for the SD-Group the magnitudes (*a*) of the facilitation expected from use of the same serial order during the five pairing-trials, and (*b*) of the decrement anticipated from use of a different order during the testing-trial.

Prediction 1 was based on the expectation that use in the stimulus-position of highly confusable, unfamiliar terms (as was the case in the present study) would lead to considerable difficulty in learning the S–R pairs, and a consequent increase in the tendency for serial position to serve as a cue.[5] The probability, then, would be increased that the S–R connection which gets established would involve the cue for serial position (*e.g.* the number of the serial position) and the response-term. During the test of retention there would thus be a tendency for responses to be emitted in the order corresponding to that in which their stimulus-terms appeared during the pairing-trials. For the SS-Group this would result in correct responses; for the SD-Group, in errors.

The low associative values of the nonsense-syllables and the use of a limited number of consonants from which these syllables were composed led to Pridection 2. Since any cue of serial-position would probably be more familiar than its corresponding nonsense-syllable and would be better differentiated from other cues of serial-position than was its nonsense-syllable from other nonsense-syllables in the list, and since both S-term familiarity and differentiation have been demonstrated to be positively related to rate of paired-associate learning, it was anticipated that rate of association of each response-term with its cue for serial position would be faster than with its nonsense-syllable. Thus, the number of correct responses during the test of retention was predicted to be greater for the SS- than for the Random-Group.

---

[4] McGeoch and Underwood, Tests of the two-factor theory of retroactive inhibition, *J. exp. Psychol.*, 32, 1943, 1-16.

[5] Newman, A mediation model of paired-associate learning, Office of Naval Research, Technical Report, 1961, No. 1, Contract No. Nonr 486(08).

METHOD

Sixty men, students from undergraduate courses in psychology at North Carolina State College, served as *Ss*. They were randomly assigned to an SS-, SD-, or Random-Group ($N = 20$ in each group). For half of the *Ss* from the SS- and SD-Groups the pairs were presented in Order A on the five pairing-trials, for the other half in Order B. For all *Ss* of the Random Group the pairs were presented in a different order on each pairing-trial. Half of the *Ss* in each of the three groups were tested with Order A, the other half with Order B.

Following the reading aloud by *E* of instructions for paired-associate learning, in which it was suggested that *S* not attend to the order in which S–R pairs appeared during the pairing-trials, a list of 10 S–R pairs was presented for five pairing-trials. The S-terms were nonsense-syllables ranging in associative value from 0% to 20%, the consonants of which were drawn from a pool comprised of X, V, F, and Z. The R-terms were 10 of the 12 terms of highest *'m'* value from Noble's list of dissyllabic nouns.[6] During pairing-trials, S–R pairs were presented at a 2-sec. rate on a Lafayette memory-drum, a 4-sec. interval separating pairs of trials. After the fifth trial, instructions for the test of retention were read to *S*, and he was then tested for retention. Since preliminary data suggested that a fast rate of presentation during the test-trial would result in few responses' being emitted, a 10-sec. rate was used in this study. *S* was instructed to respond with the appropriate R-term as its S-term was presented. Following the test-trial, *S* was given 60 sec. to emit all of the R-terms he could recall and was then dismissed.

RESULTS

Analyses of variance were carried out for the following dependent variables: number of correct responses on the retention-trial, number of different intralist responses (*i.e.* response-terms from the list) on the retention-trial, and number of different intralist responses on the test of recall for R-terms. The means for each of the six groups on each measure are presented in Table I. Bartlett's Chi-square for each distribution was found to be not significant ($p < 0.01$, $df = 5$). Since there was high correspondence between the data obtained with Orders A and B, their data were combined for each of the three experimental groups.

---

[6] J. A. Glaze, The association value of nonsense syllables, *J. genet. Psychol.*, 35, 1928, 255-267; C. E. Noble, An analysis of meaning, *Psychol. Rev.*, 59, 1952, 421-430.

The analysis for number of correct responses gave an F of 6.30 for Treatments ($p < 0.01$, $df = 2/57$). Use of the Newman-Keuls test of sequential-range for evaluating the significance of the difference between means of the SS-, SD- and Random-Groups revealed that the *t*s for the comparisons of SS- vs. Random-Groups and SD- vs. Random-Groups were significant ($p < 0.01$), no statistically significant difference being observed between the SS- and SD-Groups ($p > 0.05$). Thus the data for this measure of response support Prediction 2 that performance of the SS-Group will exceed that of the Random-Group, though not Prediction 1 that the performance of the SS-Group will exceed that of the SD-Group.

Since the possibility existed that serial learning might facilitate the learning of S–R pairs through facilitating R-term learning, analyses of degree of R-term learning were performed, with number of different intralist R-terms occurring both during the retention-trial and during R-term recall as dependent variables. In neither analysis was the F for Treatments significant ($p > 0.05$), data on both measures of response suggesting that rate of R-term learning did not differ as a function of order of presentation either during pairing- or during testing-trials. On each of these variables, however, the means for the SS- and SD-Groups did exceed that for the Random Group, the mean for the SD-Group on R-term recall being significantly higher than that for the Random Group on Test B ($p < 0.05$).

TABLE I

SUMMARY OF THE THREE RESPONSE-MEASURES FOR EACH GROUP

Groups	S–R-test				R-term recall	
	Correct responses		Different responses		*M*	*SD*
	M	*SD*	*M*	*SD*		
SS	2.70	1.85	5.20	1.33	5.15	1.27
SD	2.95	1.72	4.90	1.38	5.15	1.35
Random	1.30	.90	4.55	1.57	4.40	1.32

For *S*s of the SD-Group, a count was made of the number of times each response-term which had appeared in a particular serial position during pairing-trials occurred as a response to the S-term appearing on that same serial position on the test of retention. Similar counts involving the same pairs of response-terms were made for the SS-Group. If serial position served as the only cue to which an R-term had become associated during pairing-trials, then the number of such intrusions should have been greater for the SD-Group. The number of such intrusions was found to be 13 for the SD-Group and 12 for the SS-Group, suggesting that serial position did not function as the only cue for *S*s of the SD-Group.

DISCUSSION

The results of this study support the generalization that presentation of a list of S–R pairs in a constant order on each pairing-trial leads to more correct responses on a test of retention than presentation in a changing order, and is independent of the order used in the test. Whether this result is due to facilitation in learning the S–R pairs, or facilitation of their recall cannot be determined from the present data. Comparisons of the SS- and SD-Groups on number of correct responses emitted during the test of recall and on number of errors in serial-position make it clear that facilitation did not occur as a function of (a) *S*'s responding to the cue for serial position instead of to one or more of the cues available in the nonsense-syllable, or (b) the establishment of between-pair remote associations through serial presentation during the pairing trials. Further, the data for different intralist responses suggest that the facilitation in learning S–R pairs by the SS- and SD-Groups was not completely due to faster response-learning.

At least two possibilities exist for explaining the greater number of correct responses as a function of serial presentation during pairing-trials. McGeoch and Irion suggest that the advantages due to serial presentation may result from the more stable set or expectancy engendered by serial presentation.[7] The precise mechanisms by which such facilitation may be produced, however, have yet to be worked out.

A second explanation stressing *S*'s use of cues of serial-position is based on the assumptions that (a) serial presentation of S–R pairs during pairing-trials leads to the association of each R-term with a stimulus-compound composed of one (or more) of the cues available in the S-term plus the cue for serial position, and that (b) the probability that any R-term of an S–R pair will be elicited by such a stimulus-compound is greater than the probability that it will be elicited by any one of the elements of the compound.

The equivalence in test-trial performance of the SD- and SS-Groups suggests that facilitation in recall will occur whether or not the serial-position cue is presented with its S-term during the recall-test. Since the explanation proposed in the preceding paragraph assumes that facilitation in recall is due to the presence during recall of the cue of serial position and the S-term, some association involving the S-term and the cue for serial position must also have been established as a function of serial presentation of S–R pairs during the pairing trials. This assumption would appear to be required in explaining facilitation for the SD-Group though not for the SS-Group. Such an analysis suggests that for the SD-Group, presentation of each S-term is followed by emission of two mediating responses, one to some or to all of the cues available in the S-term itself (as would be the case with the Random Group), the other having been elicited during earlier learning-trials by the cue for serial position. Emission by the SD-Group of a correct response during the test of retention would be dependent, then, on the presence both (a) of the stimulus-properties of a mediating response to one or more of the cues available in the presented S-term, and (b) of the stimulus-properties of a second mediating response to the S-term elicited earlier in learning by the cue for serial position.

---

[7] J. A. McGeoch and A. L. Irion, *The Psychology of Human Learning*, 1952, 485.

Although performance of the SS-Group might also be explained on this basis, a simpler explanation of facilitation for this condition proposes that a single mediating response occurs to each stimulus-compound during the test of recall.[8] Since it would probably take less time for a single mediating response than for two mediating responses to occur, performance of the SS-Group would, according to this explanation, ordinarily be expected to exceed that for the SD-Group. The use of a 10-sec. rate of presentation during recall may, however, have reduced the probability that such a difference would occur.

The results of this study suggest, finally, that cues of serial position may have a marked effect on performance in serial learning. Since some of the variables manipulated in experiments on serial learning may affect performance through their effect on the functioning of serial position as a cue, and since serial position is usually not available as a cue during paired-associate learning, the divergent findings in serial and paired-associate learning may thus be partially accounted for. The identification of those variables which determine the way in which serial position functions as a cue appear to be deserving of detailed study.

SUMMARY

This experiment studied the role of serial position as a cue. Each of

---

[8] The difference in performance between the SS- and Random-Groups can be explained also using the assumptions on which Prediction 2 is based. At least one additional assumption is still required, however, to explain the equivalence of the SS- and SD-Groups.

three groups received five learning-trials during which 10 S–R pairs were presented in the same order on each trial (for SS- and SD-Groups) or in a different order on each trial (Random Group). For the SS-Group, the presentation-order for pairs during recall (Trial 6) was the same as during the pairing-trials; for the SD-Group it was not. It was predicted that performance of the SS-Group would exceed that of the SD-Group (Prediction 1) and of the Random Group (Prediction 2).

The data supported Prediction 2 though not Prediction 1. The explanation of the results assumes that (a) serial presentation of S–R pairs during pairing-trials leads to the association of each R-term with a stimulus-compound composed of one (or more) of the cues available in the S-term plus the cue for serial position, and that (b) the probability that any R-term of an S–R pair will be elicited by such a stimulus-compound is greater than the probability that it will be elicited by any one of the elements of the compound. It is suggested that divergent findings in studies of serial and paired-associate learning may be due in part to the presence of serial position as a cue during serial learning.

RATIO- AND PARTITION-JUDGMENTS

By RICHARD M. WARREN, Shimer College, and E. C. POULTON, Applied Psychology Research Unit, Cambridge, England

Stevens has claimed that for continua such as the lightness of gray papers there is a difference between the scales yielded by *ratio-* and *partition*-judgments.[1] The evidence cited in support of this claim comes, however, from experiments which differed in other respects as well.[2] In the present study, ratio- and partition-judgments of the lightness of gray papers are obtained under comparable conditions.

METHOD

Observers. The *O*s were 250 men, mainly undergraduates of Cambridge University, although a few graduate research workers and senior members of the university were used. They all were unfamiliar with psychophysical testing.

Materials. A white, a black, and some neutral gray papers, all having a dull finish with low specular reflection, were purchased from the Mitchell Colour Card Company, London. The reflectance of the white paper was 83% and that of the black paper was 3.3%. Black cotton velvet with a reflectance of 0.4% was used with some experimental groups. The reflectances of all materials were measured both with a Macbeth Illuminometer and with a Salford Electrical Instruments, Ltd., Illuminometer.

Procedure. The *O*s were tested one at a time in the Junior Combination Rooms of several of the Cambridge colleges. After reading the typewritten instructions, *O* was taken to another part of the room and seated before the stimulus-display lying on a wooden table. After *O* made his judgment of the lightness of a gray paper (thinking that only this one judgment was required), a different gray was substituted, and *O* made a second lightness-judgment. *O* was then requested not to discuss the experiment with anyone who had not already served in it. The first gray paper again was placed in position, and the next *O* brought over. A group of 50 different *O*s was used for each of the five experimental conditions.

All stimulus-patches were viewed against a rectangular background 17.5 × 11.3 in. placed with a longer side closest. In Experiment I, the background was the

---

* Received for prior publication November 6, 1961. This research was done at the Applied Psychology Research Unit of the British Medical Research Council, Cambridge, England. It was supported in part by the Air Force Office of Scientific Research of the Air Research and Development Command, United States Air Force, through its European Office, and in part by the British Medical Research Council. We are indebted to Dr. M. Stone for advice upon the statistics.

[1] S. S. Stevens, The psychophysics of sensory function, *Amer. Scientists,* 48, 1960, 232.

[2] S. S. Stevens and E. H. Galanter, Ratio scales and category scales for a dozen perceptual continua, *J. exp. Psychol.,* 54, 1957, 377-411.

standard white paper. In Experiment II, the background was the black cotton velvet. The ambient illumination always was at a level adequate for comfortable reading.

EXPERIMENT I (WHITE BACKGROUND)

Group R (ratio-judgments). *O* read the following ratio-instructions: "You will see a gray patch of paper lying on a white blotter. If the lightness of the white blotter is called 100, what (smaller) number describes the lightness of the gray patch? Please write down your answer on the slip of paper provided."

O then viewed a small rectangular gray patch 2.0 × 1.5 in. with a reflectance 22% of the white paper's which lay at the center of the white background-sheet, the long sides of the gray patch parallel to the long sides of the white paper. After the first lightness-judgment, this gray paper was replaced by a gray paper of the same size having a reflectance 53% of the white paper's, and a second lightness-judgment made.

Group P_1 (partition-judgments without black patch). *O* read the following partition-instructions: "You will see a gray patch of paper lying on a white blotter. If the lightness of the white blotter is called 100 and the lightness of black called 0, what number describes the lightness of the gray patch? Please write down your answer on the slip of paper provided."

Except for the difference in instructions, the procedure was identical with that employed for Group R.

Group P_2 (partition-judgments with black patch). The partition-instructions were identical with those of Group P_1 except that the following sentence was added at the end: "Remember, White = 100, Black = 0." The display consisted of a black patch 2.0 × 1.5 in. presented along with the gray comparison-patch, both lying on the white paper and oriented with their long sides parallel to the long sides of the white sheet. The gray and black patches were separated from each other by the same distance separating each from the nearer short side of the white paper, and both were equidistant from the long sides of the white paper. The first gray paper had a reflectance of 22% relative to white as in Groups R and P_1 and was replaced, as in the other groups, by a patch having a reflectance of 53% for the second lightness-judgment. The black patch stayed in the same position for the first and second lightness-judgments. Half the *O*s had the black patch to the right of the gray one, and half had it to the left.

EXPERIMENT II (BLACK BACKGROUND)

Group R (ratio-judgments). *O* read the following ratio-instructions: "You will see two paper patches, one white and one gray, lying on a blotter. If the lightness of white is called 100, what (smaller) number describes the lightness of the gray patch? Please write down your answer on the slip of paper provided." The blotter serving as background was black cotton velvet. The background and the stimulus-patches were the same size as those used in Experiment I. The gray and white patches were separated from each other by the same distance separating each from the nearer edge of the black background, and both were equidistant from the long edges of the background. The first gray paper had a reflectance of 22%, and was replaced, as in Experiment I, by a patch having a reflectance of 53% for the second lightness-judgment. The white patch stayed in the same position for the first and second judgments. Half the *O*s had the white patch to the right of the gray one, and half had it to the left.

Group P (partition-judgments). *O* read the following partition-instructions: "You will see two patches of paper, one white and one gray, lying on a black background. If the lightness of white is called 100 and black is called 0, what number describes the lightness of the gray patch? Please write down your answer on the slip of paper provided. White = 100." Except for the difference in instructions, the procedure was identical with that employed for Group R in Experiment II.

If any *O* did not follow instructions (*i.e.* if his responses were not between the designated limits of 0 and 100, or if he failed to give a lower number to the darker of the two grays), he was eliminated from the experiment.

TABLE I

RATIO VS. PARTITION JUDGMENTS OF GRAY PAPERS (WHITE BACKGROUND) IN EXPERIMENT I

Group	Description	Patch of 22% R (judged first)			Patch of 53% R (judged second)		
		Median	*M*	*SE*	Median	*M*	*SE*
R	Ratio	39.5	37.9	2.6	70	63.3	2.7
P_1	Partition without black patch	40	42.3	2.2	72.5	71.0	1.4
P_2	Partition with black patch	40	43.9	2.0	70	67.4	1.6

RESULTS

Ratio- and partition-judgments are summarized for Experiment I (white background) in Table I, and for Experiment II (black background) in Table II. It can be seen that for each of the experiments the judgments are similar: the medians within each experiment are in very good agreement both for the first and for the second judgments, and the corresponding differences between means in no case approach significance at the 5% level.[3] The effect of background-reflectance upon lightness is in agreement with that reported in the literature, *i.e.* the darker background causes the gray stimulus to appear lighter.[4]

DISCUSSION

Since the present study is a test for possible differences between ratio- and partition-judgments, it is essential to minimize extraneous differences which could influence the results. The separate groups of *O*s employed for ratio- and partition-judgments were drawn from the same population.

---

[3] Because of the skewed distribution of responses, a non-parametric test (the two-tailed Mann-Whitney *U*-test) was employed.

[4] S. M. Newhall, D. Nickerson, and D. B. Judd, Final report of the O.S.A. subcommittee on the spacing of the Munsell Colors, *J. opt. Soc. Amer.*, 33, 1943, 385-418.

Precautions were taken to avoid any influence of previous psychophysical experiments by employing *O*s who never had served in such studies before, and had no knowledge of psychophysical theory. Influence of prior judgments within this experiment was avoided by considering the first judgments separately.[5]

The ratio-judgments of Experiments I and II were unequivocally ratio and not partition according to the criteria set by Stevens, since the instructions called only for lightness-judgments relative to the white standard.[6] Further, the viewing conditions of Experiment I were the simplest possible for ratio-judgments—the only stimuli displayed were a white standard and a gray. The 'partitioning groups' in both experiments clearly received partitioning rather than ratio-instructions. All partition-scales, according to Stevens, "are constructed by one or another variation on the procedure that Plateau invented" when he had "artists paint a gray that seemed to lie halfway between black and white."[7] The partition-procedure in the present study also involved subjective judgments of position along the continuum between black and white (in this case numerical estimates.)

TABLE II

RATIO VS. PARTITION JUDGMENTS OF GRAY PAPERS (BLACK BACKGROUND) IN EXPERIMENT II

Group	Description	Patch of 22% R (judged first)			Patch of 53% R (judged second)		
		Median	*M*	*SE*	Median	*M*	*SE*
R	Ratio	55	51.5	2.4	77.5	74.7	1.9
P	Partition	60	55.0	2.7	80	75.7	2.2

In Experiment I, no black standard was visible for partitioning Group P_1, and it might be argued that, instead of following instructions, these *O*s made ratio-judgments relative to the white standard alone. An additional partitioning group (P_2) therefore viewed both white and black standards as well as the gray stimulus. As shown in Table I, the judgments of the three experimental groups (Groups *R*, P_1, and P_2) were equivalent (not only for the first gray, but for the second as well). Thus, for Experiment I, a distinction between ratio- and partition-judgments is unjustified.

In Experiment II, a black background was employed rather than white as in Experiment I. It is well established that a darker background causes a gray to appear lighter,[8] and such was the case in the present experiments. The important fact for the present study is that in Experiment II we find again that judgments of Group *R* (ratio) and Group *P* (partition) are very much the same, which affords further evidence for the equivalence of ratio- and partition-judgments.

---

[5] See S. S. Stevens and E. C. Poulton, The estimation of loudness by unpracticed observers, *J. exp. Psychol.*, 51, 1956, 71-78, esp. Experiment 1.

[6] Stevens, *op. cit.*, 231-234.

[7] Stevens, *ibid.*, 232.

Let us look more closely at Stevens' distinction between partition (category) and ratio- (magnitude) scales. He claims that, for more than two dozen subjective continua (including lightness of gray papers), "the magnitude scale is a power function, the discriminability (*jnd*) scale approximates a logarithmic function, and the category scale assumes a form intermediate between the other two. Over the different sense modalities, these relations among the three scales are strikingly invariant; they constitute one of the really stable aspects of psychophysics."[9] Yet power-functions have been proposed by several investigators on the basis of partition-judgments.[10] Indeed, the proposal of a psychophysical power-law by Plateau in 1872 was based upon a partition-experiment.[11] Furthermore, ratio-judgments of lightness have been reported which are in agreement with the larger body of literature on partition-judgments of lightness.[12]

Quite different results, however, were reported for ratio-judgments of lightness by Stevens and Galanter in 1957.[13] They did not obtain lightness-judgments relative to white as in previous ratio- and partition-studies, but determined judgments relative either to gray or to 'black' standards. The experimental findings of Stevens and Galanter have been confirmed in essential details by Torgerson who used similar experimental conditions.[14] These authors did not consider that the differences between the ratio-scales they obtained and the other experimental determinations of lightness might be due to differences in experimental procedures, but a recent experiment by Poulton and Simmonds indicates that the psychophysical function relating ratio-judgments and reflectance does indeed depend upon whether a white paper or a gray paper is presented as the standard lightness for evaluation of the relative lightness of the comparison-gray.[15]

---

[8] Newhall, Nickerson, Judd, *op. cit.,* 416.

[9] Stevens, *op. cit.,* 233.

[10] For a brief review of this literature, see R. M. Warren and E. C. Poulton, Basis for lightness-judgments of grays, this JOURNAL, 73, 1960, 380-387.

[11] J. A. F. Plateau, Sur la mesure des sensations physiques, et sur la loi qui lie l'intensité de ces sensations à l'intensité de la cause excitante, *Bull. Acad. Roy. Belg.,* 33, 1872, 376-388.

[12] Warren and Poulton, *op. cit.,* 380-387.

[13] Stevens and Galanter, *op. cit.,* 398-400.

[14] W. S. Torgerson, Quantitative judgment scales, in Harold Gulliksen and Samuel Messick (eds.), *Psychological Scaling: Theory and Applications,* 1960, 21-31.

There can be little doubt that different results may be obtained in partition-experiments and in ratio-experiments, but different results may be obtained from one partition-experiment to another, or from one ratio-experiment to the next. It is suggested that the characteristic differences between ratio- and partition-judgments claimed by Stevens do not exist, this erroneous distinction resulting from artifacts introduced by confounded differences in experimental procedure.

SUMMARY

The present study was designed as a direct test for the validity of Stevens' claim that psychophysical ratio- and partition-judgments are fundamentally different. It was found that these judgments are equivalent when comparable experimental conditions are used for lightness-judgments of gray papers. It is considered that an erroneous distinction between ratio- and partition-judgments has resulted from failure to consider artifacts introduced by incidental differences in experimental procedure.

---

[15] E. C. Poulton and D. C. V. Simmonds, ms. in preparation.

UTILITY AND LEVEL OF ASPIRATION

By SELWYN W. BECKER, University of Chicago, and SIDNEY SIEGEL,
Pennsylvania State University

In a recent theoretical formulation, it was suggested that the setting of a level of aspiration should be viewed as making a decision. Level of aspiration was defined as a point on an individual's scale of the utility of his goals.[1] It was also asserted that when the goals form a discrete set on a scale of achievement, then at least an ordered-metric measurement of the individual's utility of those goals is necessary to identify his level of aspiration. Furthermore, the level of aspiration is associated with the goal that bounds the top of the largest distance, *i.e.* the largest difference in utility between adjacent goals. This formulation was based on the proposition that, for an individual, there is some utility, positive or negative, associated with obtaining any goal on a scale of achievement.

Becker and Siegel supported this formulation empirically by correlating an ordered metric measure of the level of aspiration with an independent measure of the level of aspiration.[2] To supplement this correlational evidence the present paper offers an *experimental* test of the theory.

The most general conclusion that may be drawn from previous research on the level of aspiration is that success generally leads to a raising of aspiration, and failure to a lowering, but that the effects of failure are more variable.[3] For this reason, the study reported in this paper was designed to apply the proposed measure of the level of aspiration to the prediction of changes in the level of aspiration folllowing induced success or failure.

Method. At a regular meeting of an introductory course in statistics, the instructor requested volunteers to participate in a special experiment, the results of which would determine a fifth of their grade in the forthcoming mid-term examination. Of the 31 members in the class, 28 volunteered. The *S*s met in a group and each was asked to

---

* Received for publication January 12, 1959. Support of this study was provided by grants from the National Institutes of Mental Health (Grant M-1328), from the Council on Research of The Pennsylvania State University, and from a United States Public Health Service Research Fellowship (MF-6846).

[1] Sidney Siegel, Level of aspiration and decision making, *Psychol. Rev.*, 64, 1957, 253-262.

[2] S. W. Becker and Sidney Siegel, Utility of grades: Level of aspiration in a decision theory context, *J. exp. Psychol.* 55, 1958, 81-85.

[3] I. L. Child and J. W. M. Whiting, Determinants of level of aspiration: evidence from everyday life, in Howard Brand (ed.) *The Study of Personality,* 1954, 495-508; J. D. Frank, Recent studies of the level of aspiration, *Psychol. Bull.*, 38, 1941, 218-226; Kurt Lewin, Tamara Dembo, Leon Festinger, and P. S. Sears, Level of aspiration, in J. McV. Hunt (ed.) *Personality and the Behavior Disorders,* 1, 1944, 333-378.

make a number of choices (in a test booklet) in lieu of answering one question on the mid-term examination. By the use of these choices, each *S*'s 'scale of utility' of the grades A, B, C, D, and E could be discovered and, by applying the hypothesis stated above, his level of aspiration determined.[4]

When the booklets were completed, an appointment was made for later in the week when the gamble for one-fifth of his grade would occur. At the next meeting of the class, the mid-term examination was given.

At the meeting later in the week, the *S*s were told that some errors had been made in constructing the booklets and that they would, therefore, have to make new choices from a second booklet. The announcement was also made that the mid-term examination had been graded and the *S*s were given cards with letter grades representing their ostensible achievement on that test.

The procedure of assigning these 'false' grades was as follows. Actual letter grades on the mid-term examination were determined. Every *S* was assigned a grade denoting failure (grade below aspiration-level) or success (grade above aspiration-level) depending upon whether he had in fact earned a score at or somewhat above, or at or somewhat below his level of aspiration. (It was not possible to assign three of the *S*s to either group as their grades in the examination and their levels of aspiration coincided so exactly, as in the case of one *S* who aspired to a C and who performed just in the middle of the C-range of scores, that any claim that the score was above or below that level would have been obviously false.)

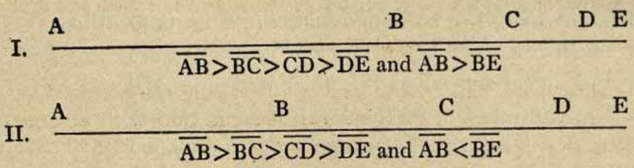

I.
A
B
C
D E
$\overline{AB}>\overline{BC}>\overline{CD}>\overline{DE}$ and $\overline{AB}>\overline{BE}$
II.
A
B
C
D
E
$\overline{AB}>\overline{BC}>\overline{CD}>\overline{DE}$ and $\overline{AB}<\overline{BE}$

FIG. 1. TWO DIFFERENT HIGHER ORDERED-METRIC SCALES WITH THE SAME ORDERED METRIC SCALE

Of the 25 *S*s remaining, 16 were assigned a letter grade above their aspirations and 9 were assigned a letter grade below their aspirations. (Of the 16 assigned grades above their aspiration-levels, 2 withdrew from the study.) After receiving their 'false' grades on the mid-term examination, the *S*s (23 in number) completed the second test-booklets. Scales based upon these second booklets were used to measure their levels of aspiration after success or failure.[5]

Technique. To test for changes in level of aspiration, it was necessary to determine what level was represented by each scale of utility. The simple ordered metric of utility — an ordering of goals and the distances between them — defines a level of aspiration that can be ranked directly, since the level is the goal bounding the top of the largest distance. When there are five goals, 24 consistent orderedmetric scales are possible, but obviously the five goals can determine only four different levels.

---

[4] Becker and Siegel, *op. cit.,* 82; also, Siegel, A method of obtaining an ordered metric scale, *Psychometrika,* 21, 1956, 207-216.

[5] At its conclusion, the purposes, techniques and results of the experiment were discussed with the entire class. They all voluntarily agreed to accept their earned numerical grades on the mid-term examination.

Higher orderedmetric measurement — an ordering of the goals, the distances between them, and the combinations of distances — is, however, a stronger measurement; it can reflect changes in utility too small for the simple ordered metric to reflect. For example, if $\overline{AB} > \overline{BC} > \overline{CD} > \overline{DE}$, the largest distance is $\overline{AB}$, but in the ordering of the combinations of distances, $\overline{AB}$ can be either greater than or less than the distance $\overline{BE}$.

The Scale II in Fig. 1 with $\overline{AB} < \overline{BE}$ when compared to the scale with $\overline{AB} > \overline{BE}$ can be said to show greater utility of the goals which lie below A. Conceivably, changes in the same direction but of even greater magnitude would result in a scale where $\overline{AB}$ was decreased to the point where it was no longer the largest distance or $\overline{BC}$ was increased to the point where it was the largest distance. A change of this nature would then define a change in level of aspiration from A to B. This pattern of change in utility (not necessarily changes large enough to affect the largest distance) makes it possible to add the following to Siegel's definition:

> A rise in level of aspiration is characterized by an increase in utility of the goals which lie above the goal bounding the top of the largest distance, and a drop in level of aspiration is characterized by an increase in utility of the goals which lie below the goal bounding the top of the largest distance.[6]

The statement implies that a rise in level of aspiration is characterized by a decrease in utility of the goals which lie below the goal bounding the top of the largest distance, and a drop in level of aspiration is characterized by a decrease in utility of goals which lie above the goal bounding the top of the largest distance.

With five goals, it is possible to construct 114 consistent ordered scales by the use of combinations of distances. If all 114 possible scales are compared for changes in utility, they yield 28 different levels of aspiration that can be arranged in rank order. An example of the reasoning used to obtain this ranking follows.

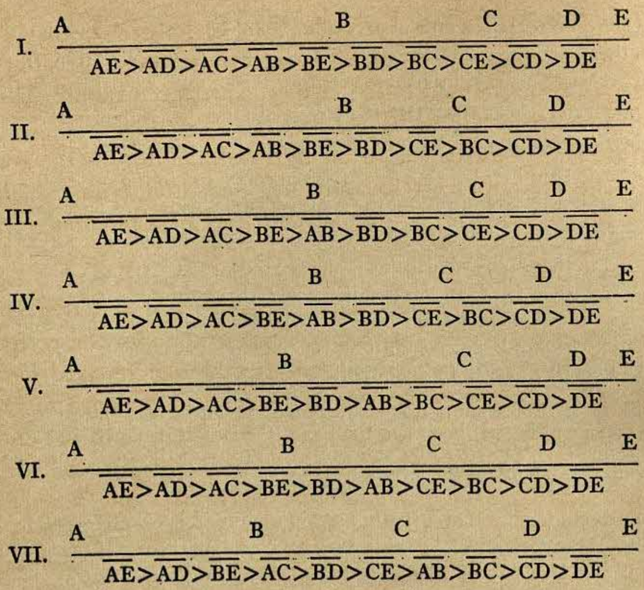
Fig. 2 shows how a pictorial representation of 7 of the 114 scales was made. In each of the seven scales, $\overline{AB}$ is the largest simple distance. Yet there are, by our definition, obvious differences in the utility of the Goal A. For example, in Scales I and II, the distance $\overline{AB}$ is greater than the distance $\overline{BE}$, while in the remaining scales the distance $\overline{AB}$ is less than the distance $\overline{BE}$. Since, in Scales III–VII, B is closer to A than to E, Scales III–VII represent a level of aspiration lower than that represented by Scales I and II. Similarly, in Scales III and IV, $\overline{AB}$ is less than $\overline{BE}$ but greater than $\overline{BD}$, while in Scales V, VI, and VII, $\overline{AB}$ is less than $\overline{BE}$ and also less than $\overline{BD}$, and therefore the latter three scales represent a lower level of aspiration than do the former two. Following reasoning of this sort, the 41 scales having $\overline{AB}$ as the largest simple distance can be put into six ranks.

A similar procedure was followed in differentiating into six ranks the 41 scales having $\overline{DE}$ as the largest simple distance.

The ranking-procedure when the largest distance is not at either end of the scale is illustrated in Fig. 3. There are 16 possible scales with $\overline{BC}$ the largest simple distance. Four of them are illustrated and they can be ranked from higher to lower levels of aspiration. The differences in utility between B and C are compared on the one hand with the difference between C and E and on the other with the difference between A

---

[6] Siegel, *op. cit.,* 253-262.



I. $\overline{AE}>\overline{AD}>\overline{AC}>\overline{AB}>\overline{BE}>\overline{BD}>\overline{BC}>\overline{CE}>\overline{CD}>\overline{DE}$

II. $\overline{AE}>\overline{AD}>\overline{AC}>\overline{AB}>\overline{BE}>\overline{BD}>\overline{CE}>\overline{BC}>\overline{CD}>\overline{DE}$

III. $\overline{AE}>\overline{AD}>\overline{AC}>\overline{BE}>\overline{AB}>\overline{BD}>\overline{BC}>\overline{CE}>\overline{CD}>\overline{DE}$

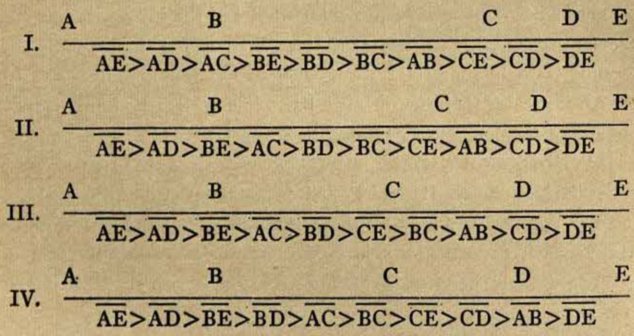
IV. $\overline{AE}>\overline{AD}>\overline{AC}>\overline{BE}>\overline{AB}>\overline{BD}>\overline{CE}>\overline{BC}>\overline{CD}>\overline{DE}$

V. $\overline{AE}>\overline{AD}>\overline{AC}>\overline{BE}>\overline{BD}>\overline{AB}>\overline{BC}>\overline{CE}>\overline{CD}>\overline{DE}$

VI. $\overline{AE}>\overline{AD}>\overline{AC}>\overline{BE}>\overline{BD}>\overline{AB}>\overline{CE}>\overline{BC}>\overline{CD}>\overline{DE}$

VII. $\overline{AE}>\overline{AD}>\overline{BE}>\overline{AC}>\overline{BD}>\overline{CE}>\overline{AB}>\overline{BC}>\overline{CD}>\overline{DE}$

FIG. 2. REPRESENTATION FOR LEVEL OF ASPIRATION RANKING OF ALL POSSIBLE CONSISTENT HIGHER ORDERED-METRIC SCALES CONTAINING ORDERED METRIC SCALE $\overline{AB}>\overline{BC}>\overline{CD}>\overline{DE}$



I. $\overline{AE}>\overline{AD}>\overline{AC}>\overline{BE}>\overline{BD}>\overline{BC}>\overline{AB}>\overline{CE}>\overline{CD}>\overline{DE}$

II. $\overline{AE}>\overline{AD}>\overline{BE}>\overline{AC}>\overline{BD}>\overline{BC}>\overline{CE}>\overline{AB}>\overline{CD}>\overline{DE}$

III. $\overline{AE}>\overline{AD}>\overline{BE}>\overline{AC}>\overline{BD}>\overline{CE}>\overline{BC}>\overline{AB}>\overline{CD}>\overline{DE}$

IV. $\overline{AE}>\overline{AD}>\overline{BE}>\overline{BD}>\overline{AC}>\overline{BC}>\overline{CE}>\overline{CD}>\overline{AB}>\overline{DE}$

FIG. 3. RANKING FOR FOUR HIGHER ORDERED METRIC SCALES WITH $\overline{BC}$ THE LARGEST SIMPLE DISTANCE

and B. The one comparison permits a ranking of the utility below C and the other of the utility above B. These changes in utility define a change in level of aspiration.

All of the 114 higher orderedmetric scales were compared in this way. When discriminations were not possible, the scales were assigned to the same rank. In all, the 114 possible scales can be assigned to 28 possible ranks.

Hypothesis. The general hypothesis to be tested was that the level of aspiration as measured by the scales of utility would rise after *S*s experienced success and would fall after they experienced failure. This formulation was operationally stated as follows:

After success there will be an increase in utility of the course grade above the previous level of aspiration, and after failure an increase in utility of the grade below the previous level of aspiration.

RESULTS

The booklets filled out by each *S* provided a pair of scales. The ranking of the scales provided a basis upon which to compare the performance of every *S* before the examination and after he had received his supposed grade. The direction of the change in the rank assigned to each of his scales was predicted from the hypothesis. Seventeen of the 23 cases showed change, and of these 11 changed in the predicted direction. By the Wilcoxon matched-pairs signed-ranks test which takes into account the magnitude of the change as well as the direction, these results were found to be significant at $p < 0.025$.[7] The data upon which the test was based are presented in Table I.

TABLE I

ACTUAL AND PREDICTED CHANGES IN LEVEL OF ASPIRATION

S	1st rank	Prediction	2nd rank	*d**	*d* ranks
A	26	up	28	−2	−2.0
B	17	up	17	0	0.0
C	26	up	28	−2	−2.0
D	6	down	1	−5	−10.0
E	6	down	9	3	6.0
F	18	up	20	−2	−2.0
G	8	up	8	0	0.0
H	26	up	6	20	17.0
I	17	down	14	−3	−6.0
J	27	up	24	3	6.0
K	26	up	9	17	16.0
L	3	down	3	0	0.0
M	14	down	24	10	15.0
N	23	up	17	6	12.5
O	28	up	28	0	0.0
P	28	up	28	0	0.0
Q	21	up	18	3	6.0
R	6	down	1	−5	−10.0
S	22	down	22	0	0.0
T	27	up	22	5	10.0
U	24	down	27	3	6.0
V	11	down	17	6	12.5
W	15	up	6	9	14.0

Σ negative ranks $= 32.0$
$p < 0.025$

* Minus sign signifies that the difference is in a direction opposite to the prediction.

---

[7] Siegel, *Nonparametric Statistics,* 1956, 75-83.

DISCUSSION

Since the evidence previously cited in support of this definition of level of aspiration was based on ordered metric measurement, it is of interest to determine whether higher ordered metric measurement provides a more sensitive measure of level of aspiration than does ordered metric measurement. The data were examined to determine how many changes had occurred on the less sensitive scales. There were only eight *S*s whose scales, based on their second performance, differed from the first (*S*s E, H, K, M, N, T, V, and W in Table I). All of these differences were in the predicted direction. There were nine more *S*s whose two scales showed a difference in the higher ordered metric but not in the ordered metric, and six of these differences were in the opposite direction from that predicted by the hypothesis. If the existing theory about level of aspiration is accepted, as it was for the purposes of this study, it appears that little information is gained through the use of the higher ordered metric rather than the ordered metric. It should be noted, however, that the use of higher ordered metric more than doubled the number of changes that were detected.

SUMMARY

Evidence is presented that a person's level of aspiration can be measured by obtaining a scale of utility of his goals. A conceptual scheme for ranking all possible higher orderedmetric scales in terms of a definition of level of aspiration is detailed and tested by predicting changes in the level of aspiration. Utility of grades was obtained before and after a midterm examination and changes in the level of aspiration were predicted.

A FURTHER STUDY OF THE INFLUENCE OF SYNTACTICAL STRUCTURE ON LEARNING

By WILLIAM EPSTEIN, University of Kansas

In a recent number of this JOURNAL an experiment was reported which demonstrated that syntactical structure facilitates verbal learning apart from the ordinarily concomitant contributions of meaningfulness, familiarity, and transitional dependency.[1] It was shown that syntactically structured material was learned more easily than matched, unstructured material when the other factors mentioned above were absent or equated. In this study the methods of complete presentation and immediate reproduction were used. Each 'sentence' was presented to *S* as a horizontally arranged series of items in the manner in which written material is usually found in English usage. The *S* read the entire sentence and then attempted to reproduce it in writing. Fewer trials were required for one perfect reproduction of the structured material than were needed for the matched, unstructured material.

The main objective of the present paper is to report an investigation of the influence of syntactical structure in serial learning. While findings bearing on this question are of interest in their own right, they are important also in another regard. The results may help to indicate the nature of the processes which underlie the original effect.[2]

EXPERIMENT I

A review of the literature failed to disclose any studies other than the original report mentioned above which deal with syntactical structure as an isolated variable. For this reason it is important to repeat the original experiment with new material prior to any further investigation.

Material. For this purpose, four categories of material, each consisting of two 'sentences,' were composed.[3] Category 1 contains sentences made up of nonsense-

---

*Received for publication June 30, 1960. This study was supported by Grant M-3600 from the US Public Health Service.

[1] William Epstein, The influence of syntactical structure on learning, this JOURNAL, 74, 1961, 80-85.

[2] A description of three possible explanations of the influence of syntactical structure may be found in the earlier paper. See Epstein, *op. cit.*, 83ff.

[3] In the earlier study, two additional categories were composed. They have been eliminated here because they are only different approaches to the same critical comparison.

syllables in combination with two functional words which have no referential meaning, *e.g.* articles, conjunctions, and prepositions. Grammatical tags, such as *ed* on verbs in the past tense and *s* on plural nouns, are appended to the syllabic stems in such a way as to simulate the requirements of English syntax. This procedure produces a series of nonsense-syllables which is readily perceived to be grammatically structured. Category 2 consists of the same material, but the items are arranged in a random, unstructured order.

In these two categories, meaningfulness, familarity, and relationships of transitional dependency have been eliminated. Only the presence or absence of syntactical structure distinguishes the two categories.[4]

The two remaining categories consist of meaningful words. These words are arranged, however, in an order which is *sententially* meaningless and within which there exists a very low level of transitional dependency between adjacent words. Category 3 contains two series of words so ordered as to meet the demands of syntactical structure. Category 4 contains the same words in a random, unstructured order.

All of these materials are reproduced in Table I.

Subjects. The *S*s were 48 students in an introductory course in psychology. Many of these *S*s had served in some type of experiment on verbal learning prior to their present participation. The *S*s were assigned to one of the four categories in alternat-

TABLE I

LEARNING MATERIAL FOR EXPERIMENT I

Category	Sentence
1*	(1) A haky deebs reciled the dison tofently um flutest pav.
	(2) The glers un cligs wur vasing un seping a rad moovly.
2	(1) deebs haky the um flutest reciled pav a tofently dison
	(2) cligs seping a wur rad un moovly glers the un vasing
3*	(1) Wavy books worked singing clouds to empty slow lamps.
	(2) Helping walls met eating trees from noisy poor lines.
4	(1) worked clouds slow empty to wavy singing books lamps
	(2) noisy trees walls from lines helping eating poor met

* Note that the 'sentences' in this category begin with a capitalized item and end with a period.

TABLE II

SUMMARY OF RESULTS OF EXPERIMENT I
(Both sentences combined for each category)

Measure of learning	Category 1		Category 2		Category 3		Category 4	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Trials to criterion	7.29	2.87	8.87	3.37	2.62	1.12	3.78	1.87
Errors to criterion	23.95	9.89	35.54	9.48	4.75	2.72	10.00	3.74

---

[4] An additional difference between the matched sentences is the order in which the items appear. It is possible to manipulate syntactical structure without altering the order of the items. This is achieved by omitting the grammatical tags, retaining only the syllabic stems in their original positions. Despite the fact that this procedure reduces the absolute amount of material to be learned, a series of this kind requires significantly more trials to be learned than a matched, structured series and is not acquired more rapidly than matched, unstructured (scrambled) material. See Epstein, *op. cit.,* 32.

ing order of their appearance in the laboratory. Each group of 12 *S*s learned the material in one of the four categories.

Procedure. The procedure was identical with that followed in the original study.[5] Each *S* performed individually under instructions to "learn all the syllables (words) in the series in the order in which they are arranged." The learning and scoring criteria were the same as those employed in the earlier study.

Results. The data are summarized in Table II. Since the results for the two sentences in each category were very similar, the data have been combined. An inspection of Table II shows that the original findings have been supported in this study. Syntactically structured material was more rapidly acquired than matched, unstructured material.[6] The results of four t-tests confirmed this conclusion. The differences in mean number of trials to criterion were significant (Category 1–2, $t = 1.57$, $p < 0.05$, $df = 46$; Category 3–4, $t = 2.46$, $p < 0.02$, $df = 46$), as were the differences in mean number of errors to criterion (Category 1–2, $t = 2.93$, $p < 0.01$, $df = 46$; Category 3–4, $t = 6.54$, $p < 0.001$, $df = 46$).

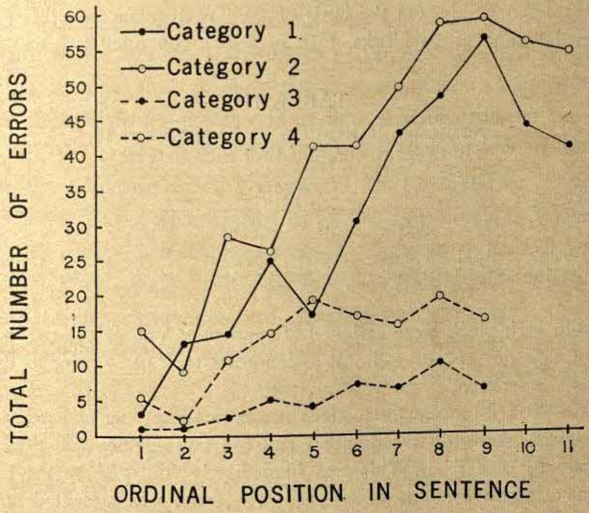


FIG. 1. THE MEAN TOTAL NUMBER OF ERRORS AT EACH POSITION FOR BOTH SENTENCES IN EVERY CATEGORY

Fig. 1 shows the errors in serial-position during learning. These curves are very similar to those obtained in the earlier study. Two aspects of the distributions will be immediately apparent: (1) none of the curves closely resembles the classical bowed serial-position curve; and (2) the shapes of the curves for the structured and unstructured material are very similar.[7]

---

[5] Epstein, *op. cit.*, 82.

[6] In one respect the present findings do not support the earlier study. In that study, the unstructured series of real words was *not* learned more easily than the structured nonsense-series. This is a most suggestive finding which has not been confirmed in the present experiment. This failure does not, however, bear on the main comparisons.

EXPERIMENT II

Experiment II is an investigation of serial learning. For this purpose the material contained in Table I was exposed on a memory-drum, and each 'sentence' was learned as a series of individual items which were presented in rapid succession.

Subjects. The *S*s were 48 students from the same population as the *S*s in Experiment I. All the *S*s had participated in a serial-learning experiment several weeks prior to the present study. On the basis of their performance on this prior task, *S*s were assigned to one of 24 blocks, each of which contained two matched *S*s. The two *S*s within each block were assigned at random to one of the two matched categories, *i.e.* Categories 1–2 or 3–4. All assignments took place prior to the beginning of the present experiment.

TABLE III

SUMMARY OF RESULTS OF EXPERIMENT II
(Both series combined for each category)

Measure of learning	Category 1		Category 2		Category 3		Category 4	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Trials to criterion	9.41	1.87	8.85	3.27	5.55	1.59	5.08	1.33
Errors to criterion	38.70	8.89	36.20	11.61	17.70	6.79	14.08	4.35

Procedure. The customary instructions for anticipatory serial learning were read to *S*. A warm-up series of five numbers and five color-names was presented on the memory-drum for five trials. This was immediately followed by the first series in the category to which *S* had been assigned. The series was presented at a 2-sec. rate with a 6-sec. inter-trial interval. The *S* was required to pronounce each item aloud. Half of the *S*s assigned to each category learned Sentence 1 first, and the other half learned Sentence 2 first. A 1-min. interval was allowed between the completion of the last trial on the first series and the beginning of work on the second series.

Results. The combined results for both series in each category are presented in Table III. Two measures of learning are recorded: (1) the mean number of trials to one errorless trial; and (2) the mean number of errors to criterion. The two-tailed *t*-test ($df. = 11$) was used to analyze the significance of the difference between the means of the two pairs of matched material for both learning measures. None of the obtained *t*s approached the value of 1.79, which is required for significance at the 5% level.[8] The errors in all four categories were distributed according to the classical bowed serial-position curve.

---

[7] A recent study of the free recall of connected material showed that the free recall of organized material is markedly different from that of unorganized material. The curve of serial position for the former material is reported to be very much like the bowed curve of serial position obtained by the method of anticipation. (See James Deese and R. A. Kaufman, Serial effects in recall of unorganized and sequentially organized verbal material, *J. exp. Psychol.,* 54, 1957, 180-187.) The apparent discrepancy between the results reported by Deese and Kaufman and the present findings may be due to the differences which exist between the respective studies in the methods employed for varying organization.

DISCUSSION

The outcome of Experiment I confirms the original study in all important respects. The similarity between the error-curves for the structured and unstructured material (see Fig. 1) suggests that the differences in speed of acquisition are not due to differences in the acquisition-techniques which *S*s employ in learning the different types of material. It is reasonable to assume that if such differences existed, they would be reflected at least partially as differences in error-distributions.

In view of the many differences between Experiments I and II, any inferences derived from a comparison of the two studies must be considered cautiously. One conclusion, however, is indicated. The absence of any significant effects in Experiment II suggests that the influence of syntactical structure which was observed in Experiment I is not based on the prior existence or experimental development, of chains of immediate probabilistic associations within the structured sentences. If such associational linkages were critical, then similar or more pronounced effects would have been obtained with the method of anticipation in Experiment II.

Instead, the effect seems to depend on the structural character of the series perceived as a unit, *i.e.* generalized linguistic form rather than sequential or semantic associations.[9]

The study of syntactical structure has rarely been attempted. Very little interest has been shown in discovering how syntactical organization is initially acquired and how this organization affects verbal performances in general and verbal learning in particular. Apart from the direct rewards to be gained from investigations of syntactical structure, there is also the possibility that the neglect of syntactical factors may lead to misinterpretations in the analysis of different but related questions. As an illustration, consider the experiments which purport to demonstrate that learning and recall of verbal passages are related to the degree of 'contextual constraint' within the passage.[10] In view of the findings reported in this paper, these demonstrations are somewhat ambiguous. The ambiguity derives from the fact that an increasing 'order of approximation' is almost always accompanied by increasing syntactical organization, which latter has been observed to facilitate acquisition apart from the contributions of statistical dependency. Is the enhanced performance reported in these experiments the result of higher-order statistical dependencies or higher-level syntactical structuring? It is not possible to answer this question since the material used in these studies involved both factors.

---

[8] Of course, the differences between the series of meaningful and of nonsense-items were significant, but this is not relevant to the objective of the study.

[9] A number of earlier writers have commented that associationistic principles may not account completely for the effects of syntactical structure. See Karl Lashley, The problem of serial order in behavior, in L. A. Jeffress (ed.), *Cerebral Mechanisms in Behavior,* 1951; James Deese, *The Psychology of Learning,* 2nd ed., 1958, 331. Similar reasons exist for questioning Osgood's interpretation which is phrased in terms of "predictive integrations" in the nervous system which match the structure of language, spanning larger linguistic units. See C. E. Osgood, A behavioristic analysis of perception and language as cognitive phenomena, in J. S. Bruner, *et al.* (ed.), *Contemporary Approaches to Cognition,* 1957, 75-118.

The introduction of syntactical factors may explain the curious finding that learning and retention tend toward the asymptope from the fourth order on through the higher orders of approximation and textual material. While passages beyond the fourth order of approximation may be at a higher level of contextual constraint in terms of uncertainty-values, they seem to be equally well structured in a syntactical sense. The facilitating effects of syntactical structure which manifest themselves equally for all passages above the fourth order may mask whatever additional benefits may accrue from increasing the degree of contextual constraint.[11]

SUMMARY

Two experiments dealing with the influence of syntactical structure on learning were reported. In Experiment I an earlier study was repeated with new material. The results confirmed the earlier conclusion that syntactical structure facilitates learning separately from meaningfulness, familiarity, and transitional dependency.

Experiment II was an invesigation of the effects of structure in serial learning. No difference was obtained between structured material and matched, unstructured material.

---

[10] See, for example, G. A. Miller and J. A. Selfridge, Verbal context and the recall of meaningful material, this JOURNAL, 63, 1950, 176-185; H. C. Sharp, Effect of contextual constraint upon recall of verbal passages, this JOURNAL, 71, 1958, 568-572.

[11] One might wish to dismiss this entire analysis by pointing out that there is no way to specify quantitatively levels of syntactical organization. Even if this objection is accepted, the analysis remains as an illustration of the possible errors which may result from overlooking syntactical structure.

DRIVE, DIRECTION OF ROTATION, AND MASSING OF PRACTICE AS DETERMINANTS OF THE DURATION OF THE AFTER-EFFECTS FROM THE ROTATING SPIRAL

By H. J. EYSENCK, R. A. WILLETT, and PATRICK SLATER,
University of London, England

Recent interest in the duration of after-effects following the inspection of a rotating spiral has been stimulated in part by the relationship this phenomenon appears to show to brain damage,[1] personality,[2] old age,[3] and drug effects.[4] A theoretical explanation of these relationships has been given by Eysenck.[5] Fundamental work on the determinants of the phenomenon itself has been rather neglected; according to Holland length of stimulation is the only major factor determining duration of after-effect of those examined.[6] This relationship is linear over much of its course.[7] More recently, investigations have been reported on the relevance to spiral after-effect (*SAE*) of drive or motivation,[8] the massing of trials,[9] and the direction of rotation of the spiral.[10] It is the purpose of this paper further to investigate these three variables and their effect on the *SAE*.

---

* Received for publication September 26, 1960, from the Institute of Psychiatry, Maudsley Hospital. We are indebted to the Society for the Investigation of Human Ecology for a grant which made this study possible.

[1] T. H. Blau and R. E. Schaffer, The spiral after-effect test (*SAET*) as a predictor of normal and abnormal electroencephalographic records in children, *J. clin. Psychol.*, 24, 1960, 35-42.

[2] H. J. Eysenck (ed.), *Experiments in Personality,* 1960, I, 205-215; II, 124.

[3] Stephen Griew and Richard Lynn, Construct "reactive inhibition" in the interpretation of age changes in performance, *Nature,* 186, 1960, 182.

[4] Eysenck, Objective psychological tests and the assessment of drug effects, *Internal. Rev. Neurobiol.*, 2, 1960, 333-384.

[5] Eysenck, *Dynamics of Anxiety and Hysteria,* 1957, 223-249.

[6] H. C. Holland, Some determinants of seen after movement in the Archimedes spiral, *Acta Psychol.*, 14, 1958, 215-222.

[7] H. C. Holland and H. J. Eysenck, Spiral after-effect as a function of length of stimulation, *Percept. Mot. Skills,* 11, 1960, 228.

[8] H. J. Eysenck and H. C. Holland, Length of spiral after-effect as a function of drive, *ibid.*, 11, 1960, 129-130.

[9] H. J. Eysenck and S. B. G. Eysenck, Reminiscence on the spiral after-effect as a function of length of rest-pause and of number of pre-test trials, *ibid.*, 10, 1960, 93 f.

[10] H. H. Spitz, Some parameters in the perception of the spiral after-effect, *ibid.*, 9, 1959, 81; C. G. Costello, Further observations on the spiral after-effect, *ibid.*, 11, 1960, 324; Effects of massed practice on the spiral after-effect and the homeostatic nature of excitation-inhibition, *ibid.*, 12, 1961, 11-14.

The following predictions were made, based partly on theory, and partly on results previously reported from this laboratory. (1) Increasing drive or motivation reduces the length of the *SAE.* (2) Massing of trials reduces the length of the *SAE.* (3) Rotation of the spiral so as to give a *contracting* effect and an *expanding* after-effect produces longer *SAE* than does rotation giving an *expanding* effect and a *contracting* after-effect. (For the sake of brevity, and as our interest is in the after-effect, we shall refer to these two forms of the phenomenon as 'expanding spiral' and 'contracting spiral,' respectively.)

Method. The spiral used has been described and illustrated elsewhere.[11] It was rotated at a speed of 100 r.p.m., at a distance of 9 ft. from *O.* Illumination was daylight, but never direct sunlight; this variable was not controlled very carefully as Holland has shown that quite large variations in illumination do not affect the *SAE* to any marked degree.[12] The *O*s were made acquainted with the phenomenon during a practice trial; they were instructed to signal the moment when the after-effect ceased to be apparent to them. Eight 30-sec. presentations of the spiral were made, each followed by a determination of the *SAE;* these presentations were *massed* in the sense that each presentation followed immediately upon the cessation of the *SAE* following the preceding presentation. After the eight massed presentations a 5-min. rest-period was introduced, followed by two further massed presentations. It was the purpose of the rest-period to allow reactive inhibition or satiation to dissipate, so that some form of reminiscence would be observable after the rest-period.[13] The assumption is made that reactive inhibition and satiation are fundamentally alike, and develop centrally and without any muscular involvement being necessary.[14]

Observers. Four groups of 36 *O*s each were used in the experiment, being tested respectively under high or low drive, and with an expanding or contracting spiral. Drive was manipulated in the manner described by Eysenck and Maxwell in their study of reminiscence on the pursuit rotor as a measure of motivation.[15] Both groups were made up of industrial apprentices 16–18 yr. of age. The *O*s, in the case of the low-drive group, had already been accepted for training by one of the biggest industrial firms in the country; while, in the case of the high-drive group, they were being subjected to a selection procedure of which the *SAE*-test was ostensibly a part. The effectiveness of this procedure has already been demonstrated in Eysenck and Maxwell's study mentioned above.

Results. The main results of the experiment are shown in Fig. 1; *SAE*s in terms of seconds are marked on the ordinate. It will be seen that all three predictions are verified. To assess the significance of these findings, a rather complex analysis of variance was performed. The main factors producing variations are 'direction' (of rotation), 'drive' (high or low) and 'pause' (before or after). With four groups of 36 *O*s each, subjected to experimentation with a particular direction and drive, and eight trials made

---

[11] Eysenck, *op. cit.*, 1957, 163-166.

[12] Holland, *op. cit.*, 219.

[13] Eysenck and Eysenck, *op. cit.*, 93.

[14] Eysenck, *op. cit.*, 1957, 150-154.

[15] H. J. Eysenck and A. E. Maxwell, Reminiscence as a function of drive. *Brit. J. Psychol.*, 52, 1961, 43-52.

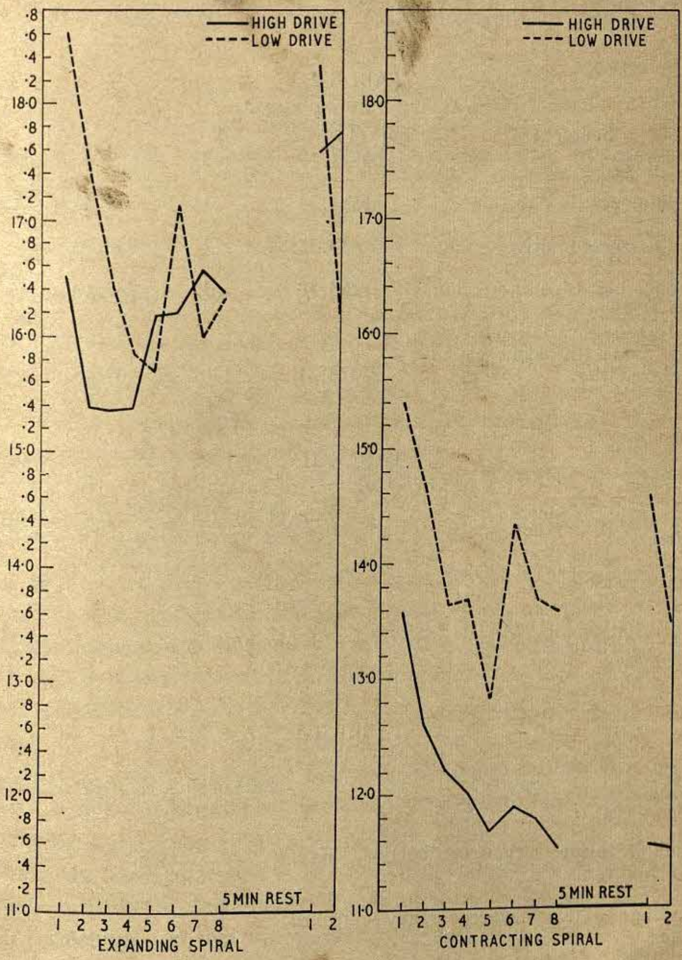


FIG. 1. SPIRAL AFTER-EFFECTS AS A FUNCTION OF DRIVE, DIRECTION OF ROTATION, AND MASSING OF PRACTICE
(Length of after-effects in seconds is shown on the ordinate; trials and rest periods are shown on the abscissa.)

before the rest-pause and two after, the total number of observations is 4 × 36 × 10 = 1440. The total observed variation about the general mean is 77,880.16 with 1439 *df*. It is separable into two parts, variation among the *O*s and variation within the *O*s. The latter part consists of each *O*'s variation about his own mean for the 10 trials, so that it has $144 \times 9 = 1296$ *df*. It can be separated again into two parts: variation among trials, with 9 *df*; and a residual, due to interaction between person and trial.

The first part of the analysis, concerning variation among the *O*s, is

TABLE I

VARIATION AMONG THE *O*S

Source	Sum of squares	*df*	*MSV*
Total observed	63821.13	143	
Due to differences in direction	5384.34	1	5384.34 ($p<0.01$)
Due to differences in drive	830.38	1	830.38
Due to direction:drive-interaction	283.10	1	283.10
Residual error, *i.e.* variation between persons under the same experimental conditions	57323.31	140	409.45

TABLE II

VARIATION AMONG THE TRIALS

Source	Sum of squares	*df*	*MSV*
Total observed	448.97	9	49.89
Due to the pause	42.80	1	42.80
Residual variation between trials not separated by the pause	406.17	8	50.77

TABLE III

RESIDUAL VARIATION

Source	Sum of squares	*df*	*MSV*
Total observed	13610.06	1287	10.58
Attributable to interactions of direction, drive, pause, and trials	278.08	27	10.30
Not associated with any experimentally controlled effects	13331.98	1260	10.58

shown in Table I. The difference between drive-levels and their interaction with direction is not significant; differences due to differences in direction of rotation are highly significant.

The second part, concerning variation among trials, is shown in Table II. The effect of the pause does not appear to be statistically significant.

Finally the residual variation within *O*s can be analyzed as shown in

Table III. It does not seem that any of the experimentally controlled effects has produced significant interactions.

There is no reason why the variation among *O*s should be comparable with the variation within *O*s, but the two parts into which the variation within *O*s has been separated are comparable with one another. We may look for some explanation of the fact that the variation among trials is significantly greater than the residual variation within persons, viz. $F = 49.89/10.58 = 4.72$ (significant beyond the 1% level).

The mean *SAE*s per trial are given in Table IV. The mean for the first trial after the pause resembles the mean for the first trial of all; both are about a unit greater than the other means. Thus we may calculate that a first trial mean = 15.682 and that the mean of subsequent trials = 14.472 and reanalyze the variance between successive trials as follows (Table V).

TABLE IV

MEAN *SAE*s PER TRIAL

	Trials before pause								Trials after pause	
	1	2	3	4	5	6	7	8	9	10
Mean	15.9	14.9	14.4	14.2	14.0	14.8	14.4	14.4	15.5	14.7

TABLE V

REANALYSIS OF THE VARIANCE BETWEEN SUCCESSIVE TRIALS

Source	Sum of squares	*df*	*MSV*
Total observed	448.97	9	
Between first and subsequent trials	337.37	1	337.37
Residual variance	111.60	8	13.95

This residual variance among trials is not significantly greater than the residual due to person × trial interactions, viz. $F = 13.95/10.58 = 1.32$. The variance, however, between first and subsequent trials is $F = 337.37/10.58 = 31.89$ (significant at 1% level). It appears that sequence-effects due to massing are present at a statistically significant level.

So far we have demonstrated a significant direction-effect ($p = 0.01$) and a significant decline-effect ($p = 0.01$); no interaction effects have reached satisfactory levels of significance, and the drive-effect has not been significant either. In view of the possible disturbing influence the massing of trials may have had on the drive-effect, it was decided to analyze just the first trial scores by themselves. The difference between the means of the high drive and the low drive groups (14.910 and 16.882 respectively) is 1.972, and the variance associated with it is 140.03. The variance among persons within groups on this trial is 43.51 (mean square with 142 *df*); hence $F = 3.218$, giving a *p*-level between 0.10 and 0.05.

While by itself clearly insufficient to prove the effect of drive, it may be said that in conjunction with the previous analogous demonstration of the depressing effect of high drive on *SAE,*[16] this finding renders it quite likely that some such effect is indeed a reality, and has to be explained by any convincing theory of the phenomenon. (On a one-tail test the effect would of course be 'significant.' Reasons for not using this test have been given elsewhere by one of us).[17]

As regards the reminiscence-phenomenon, which appears fairly clearly in Fig. 1, it also fails to receive support from the analysis. The means of the two drive-groups on Trial 8 may be substracted from those on Trial 9 to give the reminiscence-effect, *i.e.* 0.736 for the high drive group and 1.500 for the low drive group. Its significance can be considered by isolating it as a group × trial interaction with variance 10.51 and comparing it with the individual × trial interaction within groups (mean square 10.58 with 1260 *df*). *F* is approximately 1.0 and the null hypothesis is not disproved. This finding suggests that inhibition-satiation set up in the *SAE* dissipates very slowly; this may be regarded as support for Thompson's two-factor theory of inhibition that "stimulus-inhibition dissipates very slowly with the passage of time."[18]

While two of our three hypotheses have been verified at an acceptable level of significance, and the third one verified at a rather low level, it is apparent that the results are less striking than they might have been because of the very marked individual differences not associated with any of the experimentally controlled variables. This is not surprising, in view of the fact that the *SAE* has been found to be associated quite strongly not only with extraversion, but also with neuroticism.[19] It would be desirable in future experiments to include measures of these personality-variables in the experimental design in order to obtain more homogeneous groups. It is possible that such a procedure would give more impressive results.

Summary. Four groups of industrial apprentices, subdivided according to motivation and direction of rotation of spiral, were tested under conditions of massed practice for length of after-effect with a rotating spiral. It was predicted that increasing drive or motivation would reduce the

---

[16] Eysenck and Holland, *op. cit.,* 1960, 129 f.

[17] Eysenck, The concept of statistical significance and the controversy about "one-tailed" tests, *Psychol. Rep.* 67, 1960, 269-271.

[18] M. E. Thompson, A two-factor theory of inhibition. *Psychol. Rev.,* 67, 1960, 201.

[19] Eysenck, *op. cit.,* Experiments in Personality, II, 1960, 236.

length of the spiral after-effect (*SAE*), that massing of trials would reduce the length of the *SAE*, and that an expanding spiral would give longer *SAE*s than would a contracting one. All three predictions were verified, but that relating to drive failed to reach the 5% level of significance.

A FURTHER TEST OF THE EFFECT OF ISOLATION IN SERIAL LEARNING

By William A. Roberts, University of Maryland

In a recent study, Newman and Saltz investigated the effect upon serial learning of isolating an item of low meaningfulness within a list of otherwise high meaningful terms.[1] Using three different isolated items in three different experimental groups, they found that the isolated items were consistently and significantly easier to learn as responses than corresponding items in the control-list. There were significantly fewer errors made to the isolated items as stimulus-words than to corresponding control-list items although the reduction in errors was less marked than for isolated items as responses. The isolated items occurred more frequently as overt intrusion-errors than non-isolated control-items, *i.e.* there was no reduction in interference among responses. The prediction of more rapid learning of the isolated lists as a whole was also not supported.

The present study examines further the adequacy of the concepts of stimulus- and response-generalization as explanations of the effects of isolation. Previous studies, using variation in meaningfulness to produce isolation, have shown reduced errors for a single isolated item as a response. The present experiment was designed to determine whether isolation would continue to be effective when more than one item (three items) were isolated within a list. The introduction of more than one isolated item within a list should reduce the degree of isolation of each individual isolated item. The number of similar items subject to stimulus- and response-generalization is, however, still much smaller even with three isolated items than for corresponding items in a control-list. On the basis of generalization, it would still be predicted that there would be a reduction in errors for isolated items both as stimuli and responses.

The theory of generalization leads to certain other predictions concerning the effects of multiple isolation. The prediction that lists containing isolated items will be learned faster than control-lists as a result of the reduction in stimulus- and response-generalization has not been supported by either the results of Kimble and Dufort or those of Newman and Saltz.[2] One possible explanation of these negative results is that a single

---

* Received for publication January 18, 1960.

[1] S. E. Newman and Eli Saltz, Isolation effects: stimulus and response generalization as explanatory concepts, *J. exp. Psychol.*, 55, 1958, 467-472.

[2] G. A. Kimble and R. H. Dufort, Meaningfulness and isolation as factors in

isolated item does not reduce generalization-tendencies to a significant degree. If this is the case, multiple isolation may reduce generalization-tendencies to a greater degree and thus provide a more adequate test of this hypothesis.

The introduction of more than one isolated item within a list should increase response-generalization for each isolated item since similar items are available as responses. A theory of response-generalization must predict that such generalization-tendencies for isolated items will be essentially limited to other isolated items due to their similarity. It follows that a greater relative frequency of overt intrusion-errors should occur among isolated items than among corresponding items in a control-list. If there is a reduction in response-generalization among isolated and non-isolated surrounding items, it can also be predicted that a smaller relative frequency of overt intrusion-errors of isolated items into surrounding positions should occur than among corresponding items in a control-list.

METHOD AND PROCEDURE

Method: Learning materials. Four lists of 15 items each were used. Two experimental lists, I_E and II_E, and two control-lists, I_C and II_C were used. List I_C contained words of high *m*-value taken from Noble's list.[3] The words ranged from *uncle* (6.57) to *army* (9.43). *List* I_E contained the same items with the exception of Positions 5, 8, and 11 where the items *latuk* (1.26), *byssus* (1.13), and *quipson* (1.26) were placed, respectively. List II_C was made up of items of low *m*-value ranging from *neglan* (1.04) to *rennet* (1.86). List II_E was like II_C except at Positions 5, 8, 11 where *youngsters* (6.75), *army* (9.43), and *heaven* (7.91) were placed, respectively.

Procedure. A preliminary list of five nonsense-syllables was used to familiarize the *S*s with the apparatus, procedure and task. After the pre-test, S was randomly assigned to one of four groups of 15 *S*s each. Every *S* was required to learn his list to one errorless trial. The items were presented on a memory-drum at 3-sec. intervals, with an inter-trial interval of 12 sec.

Subjects. The *S*s were 60 experimentally naïve undergraduate students enrolled in the introductory course in psychology.

RESULTS

In Fig. 1 below mean errors (overt intrusion-errors plus failures to respond) are plotted as a function of serial position. The curves show no decreases in mean errors at the critical positions for List I_E. Drops at Positions 5, 8 and 11 are found for List II_E. The learning of the isolated items as responses was analyzed by converting each *S*'s errors at Positions 5, 8, and 11 into a percentage of total errors made at all positions.

---

verbal learning, *J. exp. Psychol.*, 50, 1955, 361-368; Newman and Saltz, *op. cit.*, 467-472.

[3] C. E. Noble, An analysis of meaning, *Psychol. Rev.*, 59, 1952, 421-430.

Analysis of variance of these data shows no significant difference between I_E and I_C ($F = 1.50$; $p > 0.20$), but a highly significant difference between II_E and II_C ($F = 24.65$; $p < 0.001$). The same analysis for the isolated items as stimuli (*i.e.* Positions 6, 9 and 12) reveals no significant differences between I_E and I_C or between II_E and II_C ($F < 1.00$).

It was suggested that if isolated items reduce generalization-tendencies within a serial list, Lists I_E and II_E should be learned faster than their respective control-lists. The mean total errors for Lists I_E, I_C, II_E, and II_C were 71.80, 58.13, 139.20 and 145.20, respectively. Statistical tests showed no significant differences between I_E and I_C ($F = 1.44$; $p > 0.20$) or between II_E and II_C ($F < 1.00$). Mean trials to criterion for the four lists were 12.33, 11.33, 21.73, and 21.73, respectively. The difference between I_E and I_C was not significant ($F < 1.00$). To test for reduction in response-generalization among items surrounding the isolated terms, errors made at the isolated positions both of responses and stimuli were subtracted from the total errors made at all positions in the list. The results are 33.00, 29.27, 88.13, and 71.76 mean errors for Lists I_E, I_C, II_E, and II_C respectively. Analysis of variance again showed no significant differences between I_E and I_C ($F < 1.00$) or between II_E and II_C ($F = 1.80$; $p > 0.10$).

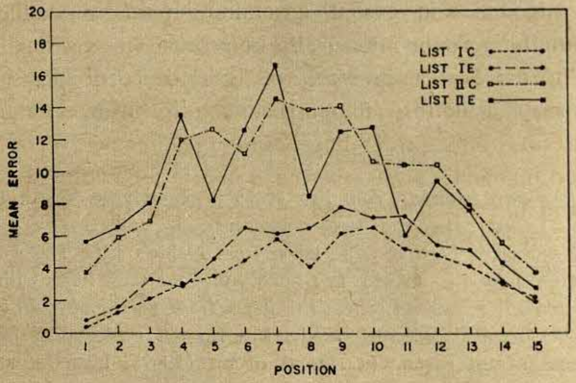


FIG. 1. MEAN ERROR AS A FUNCTION OF SERIAL POSITION

The theory of response-generalization implies that a greater relative frequency of overt intrusion-errors would occur among isolated items than among their corresponding control-items. To test this hypothesis, the intrusion-errors occurring among Positions 5, 8, and 11 (*i.e.* the number of times an item at one of these positions was given incorrectly as a response at one of the other two positions) were divided by the total intrusion-errors made at those positions to give a percentage-score. Since these scores were heavily skewed toward low percentages and in one comparison, homogeneity of variance was not met, the Mann-Whitney *U*-test was used to analyze these data. Since directional predictions were made, probabilities were determined on the basis of one-tailed tests. The differences between I_E and I_C and between II_E and II_C were in the predicted direction, the first difference almost reaching significance ($p = 0.054$), and the second reaching a higher level of significance ($p = 0.01$). If there were a reduction in response-generalization between isolated and surrounding positions, a smaller relative frequency of intrusion-errors of isolated terms into non-isolated surrounding positions should occur in Lists I_E and II_E than for items in the corresponding positions in Lists I_C and II_C. Again a percentage score was obtained by dividing total intrusion-errors made at non-isolated positions into intrusion errors made from isolated positions into surrounding positions. Differences were in the predicted direction, and *U*-tests showed a nearly significant difference between I_E and I_C ($p = 0.055$) and an insignificant difference between II_E and II_C ($p = 0.145$).

DISCUSSION

It appears that multiple isolation leads to a reduction in errors for the isolated items as responses when items of high *m*-value are isolated among items of low *m*-value, but not when items of low *m*-value are isolated among items of high *m*-value. Analysis of the responses to isolated items as stimuli shows no reduction in errors in either case. These results do not agree with the theory of stimulus- and response-generalization since this theory predicts a reduction in errors when isolated items are either stimuli or responses. It must be pointed out, however, that three factors are probably interacting in a complex manner to determine the speed of learning of the isolated items. The first factor is degree of isolation. As more isolated items are placed within a list, the degree of isolation of the individual item is lessened. The second factor is the meaningfulness of the isolated item. Noble has found that difficulty of learning varies inversely with *m*-value.[4] To the extent that generalization-tendencies exist, the third factor is the strength of such tendencies among items of the same and different levels of meaningfulness. It has been assumed that generalization-tendencies decrease as *m*-value levels diverge due to differences in similarity. Since these factors are confounded in the learning of the isolated items in this study, the failure to find reductions in errors for isolated items as stimuli and responses cannot be considered as crucial evidence against a generalization-theory. For example, in List I_E, it may be that reduced isolation plus low *m*-value are conditions which together make isolated items too difficult to learn even though there is a reduction in stimulus- and response-generalization. Kimble and Dufort suggested that a reduction in errors for a single isolated item of low *m*-value represented a reversal of the difficulty-meaningfulness relationship.[5] That degree of isolation and *m*-value interact is suggested by the results for List I_E where decreasing degree of isolation seems to have restored the inverse relationship of difficulty and meaningfulness.

---

[4] C. E. Noble and D. A. McNeely, The role of meaningfulness (*m*) in paired associate verbal learning, *J. exp. Psychol.*, 53, 1957, 16-22.

The comparison of mean total errors and trials to criterion is again not a pure test since the effect of generalization-tendencies is confounded with the effect of meaningfulness. It is interesting to note that mean total errors have a perfect ordinal correlation with mean *m*-value which further points up the potency of *m*-value as a confounding variable. When errors for each list minus errors made to isolated stimulus- and response-positions are analyzed, however, the effect upon total errors of differences in *m*-value between lists is eliminated, and the hypothesis of reduced generalization between surrounding items can be fairly tested. The fact that more errors were made to surrounding positions in the experimental lists than in the control-lists does not support that hypothesis. These findings are also in line with previous studies of isolation which have found slight increases in the errors required to learn surrounding non-isolated items.

The analysis of overt intrusion errors generally supports the notion that response-generalization was greater among isolated items than among their corresponding control-items.

SUMMARY

Two experimental groups and two control-groups learned 15-item serial lists to a criterion of one errorless trial. For one experimental group (I_E), a list containing 3 items of low *m*-value isolated among 12 items of high *m*-value was used while, for the other group (II_E), a list with items of high *m*-value isolated among items of low *m*-value was used. The control-groups learned homogeneous lists of either high *m*-value (I_C) or low *m*-value (II_C). For the isolated items as responses there was a significant reduction in errors over control-items only in the case of

---

[5] Kimble and Dufort, *op. cit.*, 361-368.

List II_E. For isolated items as stimuli, there were no significant reductions in errors in either experimental group. No significant differences in total errors or trials to learn an entire list were found between either experimental group and its corresponding control-group. The percentages of intrusion-errors were higher among the isolated items than the control-items. The implications of these results for an explanation of the effects of isolation in terms of intra-serial generalization were discussed, and the possible interaction of degree of isolation and meaningfulness was pointed out.

EPIDERMAL *b*-WAVES AND CHANGES IN BASAL POTENTIALS OF THE SKIN

By ARNOLD TREHUB, IRVING TUCKER, and JANE CAZAVELAN,
VA Hospital, Northampton, Massachusetts

Since the early investigation of Forbes, it has been generally believed that the negative (*a*-wave) and the positive (*b*-wave) components of the endosomatic galvanic skin response represent two separate and concurrent potentials produced by separately reacting neurophysiological mechanisms.[1] Wilcott and his associates in their recent studies of the waveforms of skin-potentials express a similar opinion.[2] Another possible explanation of the two components may, however, be based upon a polarization-depolarization model in which the *a*-wave represents the activity of the sweat-glands which results in a rapid increase of the negativity of the normal skin-surface with respect to sub-strata, while the *b*-wave represents a passive equilibrating depolarization through the membrane of the skin (possibly the *stratum lucidum*) which is triggered by the surge in surface negativity.

If the latter view is correct, the likelihood of the occurrence of *b*-waves following stimulation should increase as the basal potential of the skin shifts in the direction of increased surface negativity. The purpose of this study was to investigate the relationship between the direction of change in basal potential of the skin and the relative frequency of the occurrence of *b*-waves.

Method. Nine adult subjects were tested. All recordings of the skin-potential were obtained with a Gross Model 5 polygraph, using a Model 5 P1 low-level DC pre-amplifier. Two types of electrodes were used: Circular silver electrodes, 9 mm. in diameter with Redux electrode paste; and zinc electrodes, approximately 9 × 9 mm. in extent, with a zinc sulfate-gum tragacanth paste. (Results were similar for both types of electrodes.) The active electrode was applied to the volar surface of the tip of the right index finger; the indifferent electrode was placed on the right ear lobe.

The stimulus used to elicit skin potential responses was an 80-db. noise produced by a buzzer which was sounded for 5 sec. and repeated at 2-min. intervals. The

---

* Received for publication December 14, 1960.

[1] T. W. Forbes, Skin potential and impedance responses with recurring shock stimulation, *Amer. J. Physiol.*, 117, 1936, 189-199.

[2] R. C. Wilcott, C. W. Darrow, and A. Siegel, Uniphasic and diphasic wave forms of the skin potential response, *J. comp. physiol. Psychol.*, 50, 1957, 217-219; R. C. Wilcott, Effects of local blood removal on skin resistance and potential, *ibid.*, 51, 1958, 295-300; On the role of the epidermis in the production of skin resistance and potential, *ibid.*, 52, 1959, 642-649.

buzzer was controlled automatically by a timing device. The basal potential of the skin immediately preceding each stimulus-induced response (B_2) was compared with the basal level at the onset of the previous response (B_1). The difference between the successive pairs of responses represented the change in the basal potential associated with the response at B_2. Thus if the recorded potential B_2 were either more negative or less positive than B_1, the change in basal level at B_2 would be classed as negative. Whenever a positive-going response was recorded with a peak which was more positive or less negative than the potential at the onset of the response, it was classed as a *b*-wave.

Results. In a total of 270 responses, 57 *b*-waves were recorded (21.1%). When all responses are divided into those associated with a basal potential change positive in direction or no change, and those associated with a basal change negative in direction, the number of responses in each category and the percentage of *b*-waves are as shown in Table 1.

The percentage of *b*-waves which was found to occur when the basal potential shifted in the negative direction is significantly greater than the percentage under the other basal conditions ($p < 0.001$).

TABLE I

CHANGE IN BASAL POTENTIAL AND INCIDENCE OF *b*-WAVES

Responses associated with basal changes	Positive or no change	Negative change
Total number	159	111
b-waves:		
Number	18	39
%	11.3	35.1

When the data were analyzed by individuals, it was found that all but one of the 9 *S*s gave a substantially greater percentage of *b*-waves associated with basal potential shifts in the negative direction (range of differences: 12.5%–60.0%). The exception gave 1% more *b*-waves with a positive-going basal potential. A Wilcoxon matched-pairs signed-ranks test of these data shows a significant positive relationship between the occurrence of *b*-waves and negative-going basal potential ($p < 0.005$).

It was observed that one *S* produced *b*-waves spontaneously. Examination of his record revealed that his spontaneous *b*-waves occurred only after sharp increases in surface negativity as shown in Fig. 1.

Discussion. The results of this study support the conclusion that the components of the skin-potential may be explained in terms of polarization and depolarization. According to this model, the *a*-wave is an active process depending upon neural innervation of sweat-glands, while the *b*-wave does not depend upon separate neural innervation, but is a passive process resulting from the breakdown of polarized membrane under a surge of surface-negative ionic tension.

If the assumptions above are valid, then the *b*-component should be markedly affected by changes in the interstitial fluids in the epidermis and the *a*-component should be relatively insensitive to such changes. Empirical findings in accord with this conjecture were obtained by Wilcott who demonstrated that exsanguination of a limb prevents the occurrence of *b*-waves but does not eliminate *a*-waves.[3] His use of the Esmarch bandage to produce exsanguination necessarily results in relative desiccation of epidermal tissue.

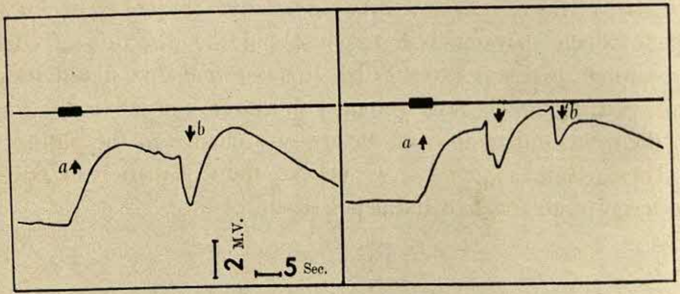


FIG. 1. SPONTANEOUS *b*-WAVES OCCURRING AFTER INCREASE IN SURFACE-NEGATIVITY

(Bar in upper left of each frame indicates onset and duration of buzzer.)

Reduction in the interstitial fluid of the epidermis might also account for the diminution of the skin-resistance response during exsanguination, since this response most likely reflects a rapid increase in conductance associated with transmembrane ion-flow.[4]

Another observation by Wilcott is pertinent. It was found that the amplitude of the *a*-waves was usually greater than normal during exsanguination.[5] This is what would be expected if an increment of polarization (an *a*-wave) were imposed upon a desiccated membrane which tends to maintain a relatively steady transmembrane potential when its fluid content is normal.

While *b*-waves almost always occur as part of a diphasic response in which the first phase is a brief negative-going wave, occasionally they are recorded without a preceding negative phase. Since the model suggested here requires an increase in surface negativity to trigger the *b*-wave, recordings which show pure *b*-responses pose a problem. A possible explanation of this relatively rare phenomenon may be that tension of polarization is so great at times that depolarization is triggered by a negative surface increment of insufficient amplitude or duration to be recorded by the usual oscillographic techniques.

---

[3] Wilcott, *op. cit.*, 51, 1958, 295-300.

[4] K. W. Goadby and H. K. Goadby, Simultaneous photographic records of the potential and resistance effects of the psycho-emotive response, *J. Physiol.*, 86, 1936, 11 P-13 P; Wilcott, *op. cit.*, 51, 1958, 295-300.

[5] Wilcott, *op. cit.*, 51, 1958, 295-300.

Summary. The relationship between relative frequency of *b*-waves in the endosomatic galvanic skin response and the direction of change in basal skin-potential was investigated. It was found that significantly more *b*-waves occurred when basal potential shifted in a negative direction than when there was either no basal change or a change in the positive direction. These results support, as we believe, the explanation of skin-potential in terms of polarization and depolarization.

PERCEPTION OF REAL MOVEMENT AS A FUNCTION OF ANGLE OF APPROACH

By DONALD R. BROWN, JAMES C. NAYLOR, and
KENNETH H. MICHELS, Purdue University

The present study was concerned with determining whether the threshold for movement would be related to angle of approach or retreat of the stimulus-object in relation to *S* where movement was of the binocular 'tracking' type. This type of real movement may be contrasted with two other types: movement of the eye relative to a stationary object; and, movement of an object in space relative to the location of the eye. It is doubtful that these all correspond to the same process. In these latter two, the probable cue is the movement of the primary stimulus across the retina, while in the first the cues are kinesthesis and the movement of stationary reference-stimuli across the retina. Data on thresholds for movement, particularly of this type, are notably sparse.[1]

Subjects. The *Ss*, four men, were graduate students in psychology.

Apparatus and procedure. The apparatus consisted of a light-source and a track along which the light could be moved. The light-source was a 6-w., 120-v., Champion lamp mounted in a 3-in. brass cylinder. The light proceeded through a hole of 0.05-in. diameter which was covered with milk-glass to provide an even gradient of intensity across the beam of light. The cylinder could be so rotated that the light was parallel to *S*'s line of sight irrespective of direction of movement. The cylinder and base were mounted on a metal-platform.

Movement of the light-source was provided by a constant-speed motor and gear-arrangement (at the rate of 3 in. per min.). The light could be moved in either direction for a distance of 6 in. by activating the motor in the desired direction of rotation. A scale (graduated in ⅛-in. units) was so mounted on the track that extent of movement could be read directly on each trial.

The entire apparatus was mounted on a table at eye-level height. The surface of the table was divided into 12 arcs radiating from its center so that the light-source could be moved along 12 combinations of direction relative to a line parallel to *S*'s line of sight. All measurements were taken from a common point, the center of a circle divided into 30°-arcs.

A plate-glass window enclosed by a 42-in. square frame was situated 29 in. in front of *S*. The 1-in. wide frame, coated with phosphorescent paint, provided a low-intensity frame of reference for the light-source.

An adaptation-light was provided by a 4½ × 5½ in. milk-glass surface located

---

* Received for publication August 5, 1960.

[1] C. H. Graham, K. E. Baker, Maressa Hecht, and V. V. Lloyd, Factors influencing thresholds for monocular movement parallax, *J. exp. Psychol.*, 38, 1948, 205-223.

168.0 in. from *S*. The intensity on the anterior surface was 44.3 ml. To mask the sound of the apparatus, *S* wore a headset through which he was stimulated by a sound of 220 ~ at an intensity of 110 db. (SPL).

The experiments were conducted in a large room in which stray light from outside sources had been eliminated. *S* was seated, without seeing the apparatus, and instructed that a light would be presented in the center of the frame of reference. He was told to report when the light was moving and the direction of movement. Direction of movement was reported in terms of clock-position. If the light was perceived as moving directly toward *S*, he responded by reporting "six o'clock," etc. If *S* reported the direction of movement incorrectly, the light continued to move; a trial was terminated when a correct judgment of direction was made. To assure a constant level of dark-adaptation, each trial was preceded by 15 sec. of light-adaptation and 45 sec. of dark-adaptation. Adjustments of the apparatus were made in total darkness in order to minimize cues which *S* could use to make judgments of movement. All trials were binocular and a headrest was provided to assure that *S* viewed the stimulus from a constant position.

TABLE I

MEANS AND STANDARD DEVIATIONS FOR ALL COMBINATIONS OF TREATMENTS

Approach or retreat	Approach				Retreat			
	Right		Left		right		left	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
0°	0.0	—	0.0	—	0.0	—	0.0	—
30°	2.62	.88	1.72	.74	5.28	2.29	5.18	2.20
60°	4.02	.52	3.18	2.01	4.17	2.23	2.83	1.40
90°	4.04	.83	5.01	.93	6.88	.02	4.50	1.60

Experimental design. The full design varied three factors: Approach *vs.* Retreat, Right *vs.* Left, and Angle of Approach or Retreat with four levels (0°, 30°, 60°, and 90°). To reduce the number of combinations of treatments for each *S*, an incomplete block-design was used. This was accomplished by breaking Factor C (angle) into two pseudo-factors, C_1 and C_2, each at two levels. Two blocks of eight combinations of treatments were used, and two *Ss* were assigned at random to each of the two blocks. The combinations of treatments were assigned to the blocks in such a way that one-third of the ABC interaction was confounded with blocks. This allowed complete within-block information on all effects except the ABC term, where only two-thirds within-block information was available. Within each block, combinations of treatments were assigned at random.

Results and discussion. The basic data used in the analysis were measures of the extent of movement of the stimulus. An analysis of variance computed on these data indicated that only the effects of Angle of Approach were significant ($p < 0.01$, $df = 3, 14$). The means and standard deviations for all combinations of treatment are summarized in Table I. Examination of these data shows that when the light was either approaching or retreating from *S* without lateral displacement from the point of origin (angle of approach or retreat = 0°), an accurate judgment of movement was not made in any instance. Under these conditions, the only cues for movement were change in size and brightness. Evidently these cues were not sufficient to judge movement under the conditions of the study. In addition, the data show that the closer the angle of approach approximated the line of sight, the more movement was necessary to obtain a correct judgment by *S*. Whether the light was retreating or approaching from either the right or left had no effect upon perception of movement.

Since the dependent variable was distance moved, and since the speed was constant regardless of the angle of approach, the major factor would seem to be rate of movement per unit of visual angle. The closer the angle to the line of sight, the less was the *rate* of movement in terms of degrees of arc per sec. This would indicate, within the limits of the rates involved in this study, that an increase in speed of a point-source of light will facilitate the perception of movement. Since other authors report a significant increase in threshold for high speeds in relation to low velocities, the present writers suggest that these results are not contradictory to previous research but that the relationship between speed and threshold is a curvilinear one; the threshold becomes lower as the speed increases up to a point beyond which it begins to rise again.

Summary. The study reports an attempt to determine the effect of angle of approach on amount of movement necessary to perceive a point-source of light as being in motion. It was found that as angle increased threshold decreased. This result was interpreted as being primarily a function of rate of movement. In the light of previous work it is suggested that the relationship between rate of movement and threshold is curvilinear.

APPARATUS

A WATERLESS THERMAL STIMULATOR

By F. NOWELL JONES, PAUL A. TWELKER, and DAVID SINGER,
University of California, Los Angeles

The best-controlled thermal stimulator so far reported is Wagoner's Thermulator.[1] This apparatus achieves a high precision of temperature-regulation and a high constant of thermal change by means of a water reservoir and a device which switches the flow from one line to another. As is the case with any device utilizing water, this apparatus is large, not easily portable, and subject to floods. The instrument described here escapes these disadvantages and it is small and not dependent on water or other liquids.[2]

Recent work on semiconductors has permitted the development of devices which utilize the "Peltier effect." As has been known for many years, the junction between two dissimilar metals will be warmed by an imposed electromotive force (*emf*) of one orientation, and cooled by an imposed *emf* of opposite orientation. Practically, this has been of very little importance because resistive heating in the metals is so great as to overcome the relatively small cooling effect at the junction. Junctions between oppositely oriented semiconductors behave as thermocouples. Furthermore, internal resistance can be made so low that the thermal effects at the junctions due to the Peltier effect are not overpowered by resistive heating; practicable devices may therefore be designed around them.[3]

For our purposes a unit developed by the Ohio Semiconductor Company proved most suitbale.[4] This device was originally designed for the cooling of transistors in electronic apparatus. Fig. 1 shows the essential parts as adapted to our purposes. The apparatus consists of two columns of bismuth telluride of *n* and *p* orientation respectively, joined into a thermocouple by a 7-mm. copper stimulating tip and the large copper plates which act also as heat sinks. A small copper-constantan thermocouple is inserted in the stimulating tip and is connected to an automatic

---

[1] D. R. Kenshalo, John Paul Nafe, and W. W. Dawson, A new method for the investigation of thermal sensitivity, *J. Psychol.*, 49, 1960, 29-41.

[2] This apparatus was developed during work on Contract No. DA-49-007-MD-1007 between the Office of the Surgeon General, U. S. Army, and the University of California.

[3] A good technical discussion and entry to the Russian literature may be found in T. S. Shilliday, Performance of composite Peltier junctions of Bi_2Te_3, *J. appl. Physics*, 28, 1957, 1035-1042.

[4] Ohio Semiconductors, Inc., 1035 W. Third Ave., Columbus 8, Ohio. The stimulator is a special modification of their TA-11 unit.

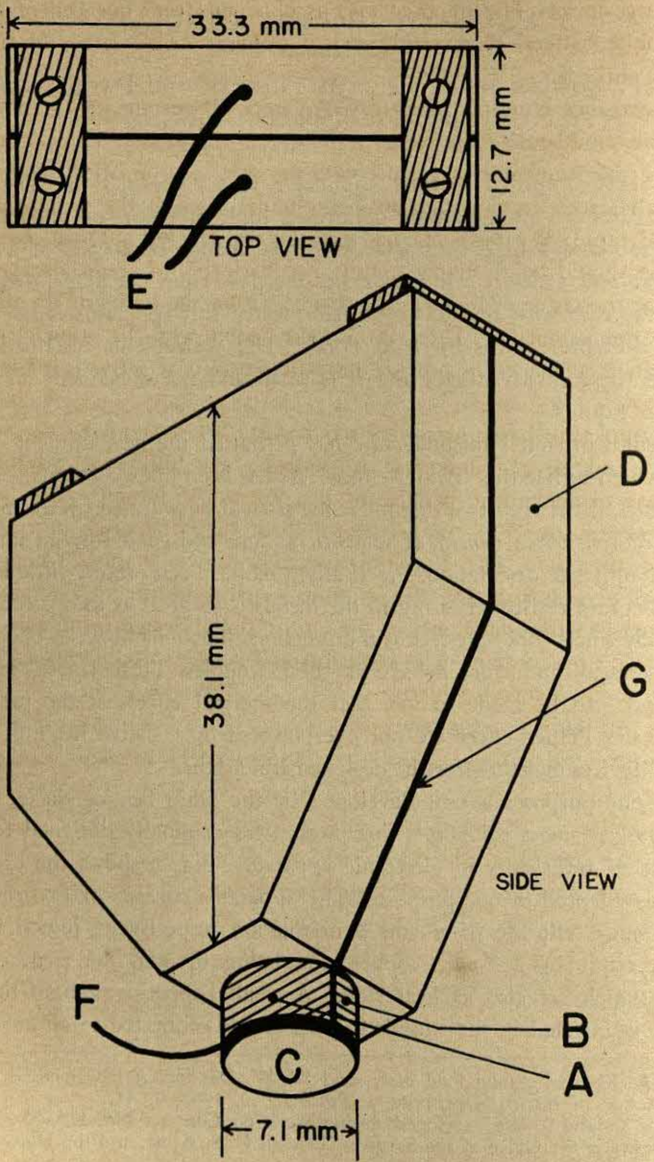


FIG. 1. THE SEMICONDUCTOR THERMOCOUPLE

The oppositely oriented bismuth telluride semiconductor A and B form a junction through copper disk C; D is one of two 'heat sinks,' separated by insulation G; while F is a thermocouple inserted in disk C. Connections are made to the copper heat sinks from a 2-v. battery through leads E.

recording device. The imposed *emf* is obtained from one cell of a 6-v. automobile battery. A reversing switch makes possible either heating or cooling at the tip.

Performance is quite satisfactory. An imposed current of 15 amp. will drop the stimulator to the freezing point. Starting at skin-temperature and keeping the stimulator in contact with the skin, a drop of 9° C. requires less than 3 sec., the time-constant being under 2 sec. A rise in temperature is more rapid. We have not had occasion to develop automatic controls but this should be relatively simple. Feedback-control could also be used to shape the curve of temperature-change within the limits of the thermal lag of the stimulator. There is also no reason why the stimulating tip should not be shaped to provide a smaller surface, if care is used to keep mass low.

The apparatus has proved very satisfactory. We have used it for several months and have found that it possesses the following advantages: (1) ease of mounting, because of the absence of tubing; (2) constant readiness for use; (3) compactness both of the stimulating unit and associated apparatus; (4) a continuous range of temperatures which, incidentally, permits resetting the stimulus-series according to *S*'s skin temperature; (5) the availability of the unit in prefabricated form; and (6) a reasonable time-constant. The versatility and convenience of this apparatus lead us to recommend it highly.

APPARATUS NOTE

THE USE OF A SUMMARY CARD-PUNCH TO GENERATE STIMULI AND TO COLLECT DATA SIMULTANEOUSLY

McConnell, Polidora, Friedman, and Meyer[1] have described an ingenious use of the IBM 526 Summary Punch for comparative psychological research. A somewhat simpler modification has been made to the IBM 526 in our laboratory that it may be used simultaneously as a general purpose generator of stimuli and as a collector of data in psychophysical experiments.

This Summary Punch has two stations. At the first, data may be read from one card and duplicated onto another card placed at the second station. At the read station, the presence of a hole in the card allows a pin-sensing contact to close and conduct a 48-v. pulse to the punch device. Thus, by leading off parallel conductors from the pin-sensing contacts to an external connector, 12 possible bits become

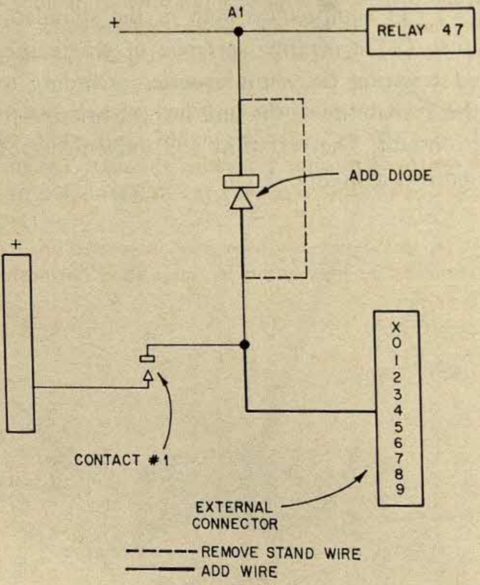


FIG. 1. DIAGRAM SHOWING THE MODIFICATION OF THE CIRCUIT OF THE SENSING-PIN
(One of 12)

---

[1] D. McConnell, V. J. Polidora, M. P. Friedman, and D. R. Meyer. Automatic reading and recording of digital data in the analysis of primative behavior. Paper read at *11th Annual Conference on Electrical Techniques in Medicine and Biology*, Minneapolis, November 1958.

available for the control of external equipment. Data may be collected from the subject directly through the summary-punch cable, or indirectly by having the experimenter key-punch the correct responses.

The difficulty in using the standard machine in this fashion lies in the fact that key-punching or entry through the summary punch-cable produces a back circuit through the pin-sensing contacts and the stimulus control-circuit is thereby activated. The circuit shown in Fig. 1 successfully avoids this difficulty and, in a simple fashion, makes the machine suitable for simultaneous presentation of a stimulus and collection of the resulting data.

To use the machine in an experiment, the stimuli are punched on a master card whose corner is clipped to activate the device that controls punching from interspersed cards. This is an accessory on the IBM 526 summary-punch. The master card is, therefore, passed directly into the read station and the punch is programmed to read as many columns of the master card as are desired. These data set the external apparatus, which probably would be a relay-tree selecting one of 4096 possible stimuli that might be selected on the basis of each of the columns read. Simultaneously, these same data are reproduced in the same columns on an unclipped card that is now in the punching-station. After these operations, the Summary-Punch is programmed to wait for other information (response) to be punched into the keyboard or entered through the summary-punch cable. The entry of the response-data completes the punching and automatically advances the next master card into the read-station and the cycle is repeated.

This device has the advantage of minimizing human intervention in the collection of the data and making available punched cards which may be conveniently analyzed on an electronic machine processing the data. The master corner-clipped cards may be used repeatedly, since only the data-cards are required for the final analysis.

In our laboratory, this device has been used for controlling random presentation of varying current-levels to nerves, but its application obviously is much wider.

IBM Research Center
Yorktown Heights, N.Y.

WILLIAM R. UTTAL

NOTES AND DISCUSSIONS

AN ARTIFACT OF THE BRUNSWIK RATIO

The index of perceptual constancy put forward by Brunswik in 1928,[1] together with its logarithmic equivalent as proposed by Thouless,[2] seems to find occasional use by investigators. It may, therefore, be worth noting that this index is subject to an algebraic defect that is at least partly responsible for a paradoxical generalization originated by Hsia and corroborated by Osgood.[3] This is to the effect that, except for the case of shape-judgments, the degree of constancy typically increases with the absolute stimulus-difference between standard- and comparison-objects. For example, Osgood reworked some of Brunswik's data on size-constancy to provide the finding that when the standard is 2 m. from the comparison series, the Brunswik ratio is 0.50; at 6 m. it is 0.78, and at 10 m. it is 0.96. As Osgood says, this is a very puzzling result. Hsia suggests that "in the adjustment to a changing situation, possibly greater effort is dispensed to overcome a greater change of conditions."[4] Osgood offers an alternative explanation which seems no more plausible.

At least part of the explanation lies in the algebraic form of the Brunswik ratio. From school algebra it is commonly known that an expression of the form

$$(a-x)/(b-x),$$

where a and b are any constants, approaches unity in the limit as x increases. This is shown by rewriting the expression in the form

$$[(a/x)-1]/[(b/x)-1]$$

and noting that as x approaches infinity, the terms a/x and b/x must approach zero. For the case of size-constancy, we may write the Brunswik ratio in the form

$$C=(R-P)/(B-P) \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad [1]$$

where R is O's estimate of size, B is the measured size, and P is the projective size. Consider an experimental arrangement, such as Brunswik's in which standard-objects are set at varying distances *Dos* from *O*, the size of the standards being judged by reference to a set of comparison-objects at a fixed distance *Doc* from *O*.[5] The comparison series is behind the farthest standard. In this case

---

[1] Egon Brunswik, Zur Entwicklung der Albedowahrnehmung, *Z. Psychol.*, 109, 1928, 68.

[2] R. H. Thouless, Phenomenal regression to the real object: I, *Brit. J. Psychol.*, 21, 1931, 339-359.

[3] Yun Hsia, Whiteness constancy as a function of difference in illumination, *Arch. Psychol.*, 39, 1943, (No. 284), 60; C. E. Osgood, *Method and Theory in Experimental Psychology*, 1953, 276f.

[4] Hsia, *op. cit.*, 60.

$$P = B \ (Doc/Dos) \ \dots\dots\dots\dots\dots \ [2]$$

thus varying as the reciprocal of the variable distance *Dos*. The absolute stimulus-difference between standard- and comparison-objects increases with the distance between them, *Dsc*. Thus the general finding reduces for this case to the assertion that the degree of constancy increases with *Dsc*. Substituting the expression for *P* in Equation [1] yields

$$C = [R - B(Doc/Dos)]/[B - B(Doc/Dos)] \ \dots\dots\dots \ [3]$$

from which with a little algebraic rearrangement, using the fact that $Dos = Doc - Dsc$, we obtain

$$C = (Doc/Dsc) - [R(Doc - Dsc)]/(B \ Dsc) \ \dots\dots\dots \ [4]$$

The absolute stimulus-difference increases as *Dsc* increases towards *Doc*. As *Dsc* approaches *Doc*, the first term of the obtained expression approaches unity. The second term approaches zero, provided that *R/B* is any ordinary function of *Dsc*. More precisely, this holds so long as *R/B* at worst approaches infinity more slowly than does the reciprocal of $(Doc - Dsc)$.

Though the details of the algebra differ, a similar result can be shown for brightness-constancy, as in Hsia's investigation.[6] Here it is easy to show that

$$C = 1 - [(R - B)/(A \ D)] \ \dots\dots\dots\dots\dots \ [5]$$

where *R* is *O*'s brightness-match, *B* is the intensity of the standard, *A* is the albedo of the standard and *D* is the difference in illumination of the standard and comparator. Hence as *D* increases, *C* approaches unity provided only that $(R - B)$ is neither a linear-increasing nor a positively accelerated function of *D*.

The one exception which Hsia found to his proposed generalization came from the data on shape-constancy in investigations by Thouless. In this case, constancy ratios like Brunswik's tend to increase to a maximum and then decrease again as the angle away from frontoparallel presentation increases. Correspondingly, in this case it can be shown that

$$C = (R - \cos \theta)/(1 - \cos \theta)$$

where *R* is *O*'s axis-ratio match and θ is the angle of tilt of the standard from frontoparallel. As θ increases, $\cos \theta$ decreases. Empirically, *R* decreases monotonically. Relative rates of decrease cause *C* to increase to a maximum, then decrease again. This would account for the failure of Hsia's generalization in the case of shape-constancy.

It is perhaps worth noting that Leibowitz has examined the algebra of the Brunswik- and Thouless-ratios with the object of finding *R* as a function of *C* and the relevant variables, whence *C* can be calculated as a parameter, on the assumption that C is independent of distance-variables and the like.[7] Since this assumption is clearly contrary to fact, Leibowitz's formulae would seem to be of limited usefulness.

---

[5] Brunswik, Thing constancy as measured by correlation coefficients, *Psychol. Rev.*, 47, 1940, 69-78.

[6] Hsia, *op. cit.*, 27-47.

A further peculiarity of the Brunswik ratio relates to the reliability of the judgment. Because of the algebraic form of the ratio, if the variance of *O*'s estimates were independent of the stimulus-difference, the variance of C would be an inverse function of stimulus-difference, approaching infinity as this difference approaches zero. Thus in a size-constancy experiment, when the standard- and the comparison-series are lined up at a common distance from *O*, the theoretical value of C is indeterminate, and its variance is virtually infinite. Clearly this need not imply a further paradox to the effect that the reliability of judgment improves with increased stimulus-difference. In fact, in investigations such as Gibson's[8] the *SD* of raw estimates increases with the distance between the standard and comparison-objects. This is not enough to offset the instability inherent in C at near-zero separations, hence we find numbers of anomalous C-values at small separations in the data of Brunswik, Hsia, and others.

University of New England
Australia
RODERICK P. MCDONALD

---

[7] H. W. Leibowitz, The relation between the Brunswik and Thouless ratios and functional relations in experimental investigations of perceived shape, size and brightness, *Percept. mot. Skills,* 6, 1956, 65-68.

[8] J. J. Gibson, *The Perception of the Visual World,* 1950, 183-187.

Elmer Augustine Kurtz Culler: 1889–1961

Elmer Augustine Kurtz Culler, Professor Emeritus of Psychology at the University of Rochester, died at the age of 71 years in Columbus, Ohio, on June 30, 1961.[1] Although serious illness began on June 12, 1961, the development of Parkinson's disease was apparent in 1942. It reached a stage in 1946 that terminated Culler's scientific career. From this time on he lived in Columbus except for the winters spent in Fort Lauderdale, Florida. His years as an invalid fortunately were happy ones, full of enjoyment and personal security. He required constant nursing and received it unstintingly together with love and devotion from his wife, Jane Pond Culler, and his foster son, L. Edward Culler, both of whom survive him.

Elmer Culler was born in Louisville, Ohio, on October 11, 1889. His family were members of the German Baptist Brethren, a religious sect that bans war, oaths, alcohol, tobacco, and worldly amusements, whose adherents wear simple, plain clothing and are known popularly as Dunkers. Culler's earliest professional ambition was the ministry. He attended Juniata College from 1906 to 1910, earning the A.B. degree in 1910, and Union Theological Seminary from 1910 to 1913, receiving the D.B. degree in 1913. The award of a Traveling Fellowship made it possible for him to undertake advanced study in Germany. He spent the academic year 1913-14 at the University of Berlin and 1914-15 at the University of Leipzig. He heard Wundt lecture at Leipzig in 1915 and this experience was a significant factor in his decision to become an experimental psychologist. This choice was also influenced by his older brother, A. J. Culler, who had studied both theology and psychology, who published research on verbal learning but made the ministry his career. Culler began graduate study in psychology at the University of Chicago in 1915-16 and continued it during the summer quarters from 1917 to 1922. Harvey Carr was his major professor and Culler received the Ph.D. in 1922 with a dissertation in the fields of temperature-sensitivity and psychophysics.

Culler was an instructor at Ohio State University from 1916 to 1920 and at the University of Wisconsin from 1920 to 1923. At Ohio State, he carried on the research for his dissertation. At Wisconsin, he completed the analysis of data and wrote the dissertation. At the same time he collaborated with W. S. Taylor in psychotherapy, and together they published a case study of William Ellery Leonard, Professor of English at Wisconsin, poet and author, who had written an autobiographical account of his phobia in *The Locomotive God.*[2] In 1923 Madison Bentley brought Culler to the University of Illinois with the rank of Associate, and there Culler continued his research on temperature-sensivity and psychophysics, much of which involved repetition, reworking, or further development of his doctoral dissertation. Caution was an outstanding characteristic of Culler's scientific work, a caution which delayed publication until the major aspects of a problem were solved experimentally, but which also resulted in finished work with few loose ends. Such was the case with Culler's studies in psychophysics. With no publication prior to 1926, this year recorded five major papers, two of them monographs, one of which included ten separate studies.[3] The year 1927 saw the publication of two additional articles in psychophysics and the virtual end of publication in this area.[4] These classical studies were widely recognized as major contributions in research and established Culler's distinction as a scientist. The University of Illinois promoted him to an assistant professorship for 1927-28 and an associate professorship in 1928. When the Society of Experimental Psychologists was formally organized in 1929, Culler became one of its charter members.

---

[1] Acknowledgment is made to Glen Finch, Edward Girden, Fred A. Mettler, and G. R. Wendt for personal communications and materials used in the preparation of this account.

The late 'twenties were years during which Culler was considering new areas of reesarch, areas in which his training and experience in quantitative methods could be used productively, and in which he could conduct programmatic research similar to that completed so successfully in psychophysics. Three abstracts of papers presented at meetings in 1928 and 1929 bear the titles, "Nature of the learning curve," "Function of cutaneous muscle in the perception of cold," and "An empirical study of zeor and negative intelligence."[5] None of these studies was published in full.[6] It was during this period that the studies of Leonard noted previously appeared in print.[7] Further work by Culler on the Leonard case resulted in a personality-classification with categories of exocentric and endocentric, similar to the more popular terms, extrovert and introvert.[8]

---

[2] W. S. Taylor and Elmer Culler, The problem of the Locomotive God, *J. abn. soc. Psychol.*, 24, 1929, 342-399; Note on "The problem of the Locomotive God," *ibid.*, 26, 1931, 340 f.

[3] Elmer Culler, On thermal sensitivity and the nature of sensory adaptation, *Brit. J. Psychol.*, 16, 1926, 193-198; Studies in psychometric theory:I–X. *Psychol. Monog.*, 35, 1926, 56-137; XI. The effect of practice on the distribution of judgments; XII. On the T-method: A procedure for finding the trend of any series of numbers, *J. exp. Psychol.*, 9, 1926, 169-194; XIII. The normal law: Its place in quantitative psychology, *ibid.*, 271-298. Thermal discrimination and Weber's law: With a theory on the nature and function of sensory adaptation, *Arch. Psychol.*, 13, 1926 (No. 163), 5-134.

[4] Culler, Accuracy of the method of constant stimuli: A reply to Dr. Urban, this JOURNAL, 38, 1927, 307-312; Studies in psychometric theory: XIV. On the probable error of the limen, *J. exp. Psychol.*, 10, 1927, 463-477.

[5] Culler, Nature of the learning curve, *Psychol. Bull.*, 25, 1928, 143 f.; Function of cutaneous muscle in the perception of cold, *Psychol. Bull.*, 26, 1928, 142 f.; An empirical study of zero and negative intelligence, *Proc. IXth Internat. Cong. Psychol.*, 133 f.

[6] There is one exception to this statement. The results of the learning curve study were included in Culler's final publication: Elmer Culler and Edward Girden, The

The development of electrophysiology (especially the work of Wever and Bray), the possibility of using conditioned responses of dogs to auditory stimuli for threshold-studies, and Culler's familiarity with psychophysics led him to consider animal hearing an area in which he could plan and conduct a program of research. Madison Bentley, then at Cornell, encouraged Culler and suggested that he spend his sabbatical year (1930-31) in Ithaca to learn the new techniques, to plan the research program, and to get it started. The year at Cornell was fruitful. Culler studied electronics, physiology, neuroanatomy and histology. He conducted experiments in electrophysiology and conditioning. He discovered that Bechterev's motor conditioning was superior to Pavlov's salivary conditioning for investigation of hearing in dogs. He concluded that a research program of the kind he envisaged could not be conducted by one person alone. It would require substantial funds and a team including graduate assistants and collaboration or support from individuals expert in electronics, neuroanatomy, and histology. Here again, Madison Bentley was helpful. As lay adviser to the Research Council of the American Otological Society, he was instrumental in the grant it made to Culler for establishing a laboratory and a program of research in animal hearing.

Back at Illinois, laboratory space for Culler was provided by the University in an old frame house, formerly a rooming house, at 106 South Goodwin in Urbana. Glen Finch and Edward Girden were appointed to research assistantships. Fred Mettler, then a graduate student in neuroanatomy at Cornell, whom Culler came to know well during his sabbatical year, agreed to spend his summers in the laboratory in collaborative research and to provide support in neuroanatomy and histology. Support in the construction and maintenance of electronic equipment was provided by the engineering staff of the University radio station and by graduate students in electrical engineering. During 1931-32 the old rooming house was transformed into a laboratory with an active research program. Manual labor on the part of the staff provided two sound-isolated laboratories, one for conditioning and one for electrophysiology, a surgery, a recovery room, and living quarters for cats and dogs.

Research was begun within the broad program already formulated. This

---

learning curve in relation to other psychometric functions, this JOURNAL, 64, 1951, 327-349.

[7] Taylor and Culler, *locc. citt.*

[8] Culler, A phobic case, *Brit. J. med. Psychol.*, 10, 1930, 46-69.

program called for study of the function of each component of the acoustic system from the cochlea through each way station to the auditory cortex by a combination of behavioral, electrophysiological and histo-anatomical techniques. It included investigation of the nature of conditioning and the neural correlates of conditioning at all levels of the nervous system. Initial projects were concerned with the function of the round window, the localization of frequency within the cochlea and the auditory nerve, with effects of X-radiation and exposure to loud tones on auditory acuity, with avoidance conditioning and extinction, higher order conditioning, conditioning in spinal and decorticate preparations, with the effects of cortical lesions upon auditory acuity and the formation and retention of conditioned responses.

Not all of these projects were immediately productive. Some remained persistent problems for a number of years until they reached eventual solution. Others were never solved to Culler's satisfaction. The fruitful projects, however, were reported at meetings and began to appear in print by 1933. The laboratory was now an established success and continued to receive major financial support from the Otological Society throughout Culler's career. The productiveness of the program resulted in increased publication year by year. Culler made reports on the research every year at meetings of the Experimentalists, the Physiological Society, the Midwestern Psychological Association, and the American Psychological Association. In a relatively short time he achieved eminence for his scientific contributions in audition and conditioning.

In 1936 he was promoted to a full professorship at Illinois. The year 1938 brought the award of the Warren Medal by the Society of Experimental Psychologists "for his work on the fundamental mechanisms of hearing and on the physiological basis of the conditioned reflex" and a new position at the University of Rochester as Professor and Chairman of the Department of Psychology. The Animal Hearing Laboratory, of which Culler was Director both at Illinois and Rochester, continued to be productive of research investigations in hearing, conditioning, and neuropsychology. Because of his illness, Culler's duties as Chairman at Rochester terminated in 1945. He became Professor Emeritus in 1954.

Culler received many professional honors and gave time to many professional activities. He served as associate or consulting editor to the following journals: *Journal of Comparative Psychology, Journal of Experimental Psychology, Journal of Psychology, Psychometrika,* and *Psychological Record*. He served as director of the American Psychological Association, the Eastern Psychological Association, the Midwestern Psychological Association, and the Psychometric Society, and was a member of numerous committees of these organizations. During 1940, he was President of the Midwestern Psychological Association and Vice-Chairman of the Division of Anthropology and Psychology of the National Research Council.

Space does not permit citation of the 70 papers that appeared as communications from the Animal Hearing Laboratory. Not all of these papers bear Culler's name, particularly those published during the Rochester years when administrative duties reduced his active participation in collection of data below the level he set for his name to appear as co-author. All of them, however, are his products. They had their origin in the broadly conceived, flexible program established in the early 'thirties involving the combined application of conditioning, electrophysiological, and histological techniques to problems of hearing and learning. Culler's major contributions may be classified under audition, conditioning, and neuropsychology. Representative of his contributions in audition are his measurement of acuity by the conditioned response technique and his determinations of the function of the round window, the acoustic value of components of the auditory system, the effect of lesions in the medial geniculate bodies upon hearing, and the frequency map of the cochlea of the guinea pig.[9] His work on conditioning covered higher order conditioning, the relation of forgetting to experimental extinction, the role of incentive in conditioning and extinction, and reciprocal facilitation of the conditioned and conditioning mechanisms.[10] In neuropsychology he completed research on conditioning in the decorticate animal, the effects of curare upon conditioning and the nervous system, the function of the extrapyramidal system, and spinal conditioning.[11] All of these papers are reports of experiments;

---

[9] Elmer Culler, Glen Finch, Edward Girden, and W. J. Brogden, Measurement of acuity by the conditioned response technique, *J. gen. Psychol.*, 12, 1935, 223; Correlation of auditory acuity with peripheral electrical response of the acoustic mechanism, *J. Psychol.*, 2, 1936, 409-419; W. J. Brogden, Edward Girden, F. A. Mettler, and Elmer Culler, Acoustic value of the several components of the auditory system in cats, *Amer. J. Physiol.*, 116, 1936, 252-261; H. W. Ades, F. A. Mettler, and E. A. Culler, Effect of lesions in the medial geniculate bodies upon hearing in the cat, 125, 1939, 15-23; Elmer Culler, J. D. Coakley, K. Lowy, and N. Gross, A revised frequency-map of the guinea-pig cochlea, this JOURNAL, 56, 1943, 475-500.

[10] Glen Finch and Elmer Culler, Higher order conditioning with constant motivation, this JOURNAL, 46, 1934, 596-602; Relation of forgetting to experimental extinction, this JOURNAL, 47, 1935, 656-662; W. J. Brogden, E. A. Lipman, and Elmer Culler, The role of incentive in conditioning, and extinction, this JOURNAL, 51, 1938, 109-117; William Eccker and Elmer Culler, Reciprocal facilitation of the conditioned and conditioning mechanisms, *J. comp. Psychol.*, 31, 1941, 223-231.

[11] Elmer Culler and F. A. Mettler, Conditioned behavior in a decorticate dog, *J. comp. Psychol.*, 18, 1934, 291-303; Edward Girden, F. A. Mettler, Glen Finch, and Elmer Culler, Conditioned responses in a decorticate dog to acoustic, thermal, and tac-

in them theoretical discussion is limited to the data at hand. Culler was primarily interested in function as discovered by experiment. His attitude toward theory is best expressed in his own words: "In behalf of the material here offered, let but one thing be said; no aspect of conditioning is included unless we have studied it in this Laboratory. I prefer not to theorize about any of these problems without first interrogating our animals."[12]

Culler also made a major scientific contribution [illegible] the training of students, a contribution that was contiguous with and an integral part of the Animal Hearing Laboratory. Approximately 20 persons earned the Ph.D. degree under Culler's guidance. He was adept in the apprentice technique of training research workers. At the beginning of his graduate study each new student started on laboratory research in collaboration with Culler and advanced students. He continued research, assuming more and more responsibility and contributing more and more to the research program of the laboratory. When he began the work for his dissertation, he usually had been a collaborator with Culler and others in several completed experiments. By the time of the Ph.D. degree, he had completed a rich and varied research experience, one that was objectively substantiated by a record of publication. Culler's teaching of graduate students was informal. It took place in the laboratory, sometimes during the actual conduct [illegible] experiments, at other times through discussion during natural breaks in [illegible]perimental work or at the close of experimental sessions. Much of [illegible] was by example. All of it was facilitated by Culler's treatment of students as peers. He made them feel free to express themselves, to make suggestions, to propose hypotheses, to question and to criticize. He encouraged a joint responsibility for the conduct of research and stimulated excitement in the process of inquiry. He fostered care and precision in collecting data, the checking on possible errors of procedure, the repeating of experiments, the criticism of one's own work, and the formulation of ideas, hypotheses, and problems into experimental designs that could be put to laboratory test. Students worked with Culler, not for him.

University of Wisconsin W. J. BROGDEN

---

tile stimulation, *J. comp. Psychol.*, 21, 1936, 367-385; Girden and Culler, Conditioned responses in curarized striate muscle in dogs, *J. comp. Psychol.*, 23, 1937, 261-274; Elmer Culler, J. D. Coakley, P. S. Shurrager, and H. W. Ades, Differential effects of curare upon higher and lower levels of the central nervous system, this JOURNAL, 52, 1939, 266-273; F. A. Mettler, H. W. Ades, E. A. Lipman, and E. A. Culler, The extrapyramidal system, *Arch. Neurol. Psychiat.*, 41, 1939, 984-995; P. S. Shurrager and Elmer Culler, Conditioning in the spinal dog, *J. exp. Psychol.*, 26, 1940, 133-159.

[12] Culler, Recent advances in some concepts of conditioning, *Psychol. Rev.*, 45, 1938, 134-153.

BOOK REVIEWS

Edited by T. A. RYAN, Cornell University

Plans and the Structure of Behavior. By GEORGE A. MILLER, EUGENE GALANTER, and KARL H. PRIBRAM. New York, Henry Holt and Company, 1960. Pp. x, 226. $5.00.

"The authors of this book believe that the plans you make are interesting and that they probably have some relation to how you actually spend your time during the day" (p. 5). As this quotation suggests, common sense plays a considerably greater part in this book than it does in much of current theoretical writing about behavior. This, in itself, is a hopeful sign, but the authors have done more than talk about everyday experiences and examples. The book is still an armchair kind of essay, but new language and new ideas are introduced to deal with the classical problems, and the book could have an important influence upon the development of future accounts of human behavior. In developing their ideas, the authors have been strongly influenced by the theory of computers, by cybernetic concepts, and by the theory of language, but concepts from these other fields are not allowed to dominate the psychological field and only provide helpful suggestions for new ways of looking at problems.

"*Plan.* Any complete description of behavior should be adequate to serve as a set of instructions, that is, it should have the characteristics of a plan that could guide [illegible] action described. When we speak of a 'plan' in these pages, however, the term will refer to a *hierarchy* of instructions, and the capitalization will indicate that this special interpretation is intended. *A Plan is any hierarchical process in the organism that can control the order in which a sequence of operations is to be performed*" (p. 16). In this definition, the key word is *hierarchical*. A 'plan' may govern broad 'strategy' or specific 'tactics.' The tactical or molecular plans are embedded within the strategical or molar plans. The smaller units are *nested* within the larger units, and not merely chained together end-to-end.

The 'reflex' or S-R unit is rejected as a possible unit for the analysis of behavior, and in its place the authors adopt the feedback loop as the unit. This loop they call the 'TOTE' unit (test, operate, test, exit). The hierarchical organization applies to TOTE units as well as to plans, because the "operate" phase may be made up of other TOTE units, etc. In addition TOTE units can be treated at varying levels of abstraction. The loop may represent actual energy flow in concrete physiological structures, or it may represent only "information" or "control" without reference to specific organic structures. This reviewer has been unable to discover the precise relation between the concept of Plan and the TOTE unit as conceived by these authors. At first they state that "planning can be thought of as constructing a list of tests to perform." In other parts of the book, however, Plan and TOTE seem to be interchangeable terms. Both are hierarchical in their structure with molecular units nested within molar structures.

The value of the book does not consist, however, of precise formulations or clear definitions of new ideas. Instead, it must be read as a stimulating example of an intellectual struggle to break away from some of the shibboleths of behavioristic thinking and to create a free and unfettered approach to some of the continuing problems of psychology.

One other major concept of these authors' approach to behavior is the *Image* (always capitalized like *Plan*). This term (taken from the book by the economist Kenneth E. Boulding, *The Image,* 1956) refers to the individual's perception, conception and knowledge of the world in which he lives and behaves. In this book the Image is largely taken for granted and the stress is upon the Plan. Nevertheless, the two cannot be completely separated, and some important aspects of behavior are related to the Image. In fact, the problem of motivation—why one plan is executed rather another—is a problem of the Image. The explanation for the choice of plan involves *values* which are part of the Image.

Following their general statement of position and definitions of terms, the authors introduce a chapter on "The Simulation of Psychological Processes." The chapter is of interest in its own right, and also as showing one of the influences upon the thinking of these authors. In addition they consider that it provides an important argument in convincing the died-in-the-wool behaviorist that "cognitive" and "purposeful" approaches to behavior can be respectably scientific because they can be simulated by machines. The direct contribution of machine theory to psychological theory is, however. of doubtful value. As the authors themselves make clear, we must have a conception of the process we wish to simulate before the simulation can be worked out. Simulation proves the workability of the conception, but neither proves its correctness nor produces the conception in the first place.

In the opinion of the reviewer, the most important chapter of the book is Chapter 4: "Values, Intentions, and the Execution of Plans." Here the authors' conception of Plan is clarified by a comparison with Lewin's ideas of intention and task. It becomes clear that the two approaches are closely related but not identical, and the authors take pains to specify differences in prediction for specific problems. The influence of computing machines creates a greater emphasis by Miller, Galanter and Pribram on the memory and retrieval of plans than is evident in Lewin's approach.

The remainder of the book consists of a discussion of various aspects of behavior which involve plans. The range is from instincts (built-in plans) to plans for the formation of plans, and including skills, language, problem-solving, and plans for remembering. The principal value of these chapters is in further illustration of the meaning of *Plan* to the authors. Little is added to the substance of these topics, since most of the discussion consists of casual observation. There are many suggestions for research which might be done, but little research backing for their point of view. Research is not needed to demonstrate the broad point that behavior is organized into hierarchical units, and that Plans are involved in all sorts of endeavors. Neither does lengthy arm-chair discussion seem necessary. The reader may therefore become impatient with the later chapters of the book, but he should not skip them, even so. There are frequent suggestions for research which might be done to spell out the detailed operations of plans, and often we find a really fresh point of view upon an old topic. One good example is the approach to verbal learning contained in the Chapter on "Plans for Remembering." Of these later chapters, the one on "Plans for Speaking" is the most useful in bringing out the full force of the notion of hierarchy of plans, and is a most effective attack upon S–R conceptions.

This book performs a valuable service and should stimulate much thought and fruitful discussion. At the same time it is disappointing in several ways. First, it is too bad that so much of the effort involved in the book is devoted to wrestling with the shades of Watson and Hull or with the "Images" of their intellectual heirs, instead of being used for making further progress with the theory itself. As a result, little more than the beginnings of the theory are sketched out. Secondly, there are important gaps in their conceptions which the authors do not seem to realize. For example, they see many important values in Tolman's cognitive approach, but they also are impressed with Guthrie's criticism that Tolman's rats are "left buried in thought." Miller, Galanter, and Pribram are "disturbed by a theoretical vacuum between cognition and action" (p. 11). The Plan is supposed to fill this gap. Yet the authors have almost nothing to say about why a particular plan is executed at a particular time instead of being stored or forgotten. In other words, the rats (or men) of Miller, Galanter, and Pribram could also be lost in the contemplation of the Image without putting any Plan into execution. It is true that once the behavior starts, it will be organized hierarchically and run off in a smooth manner, but only after the plan has been put into execution. In short, there would seem to be just as great a "theoretical vacuum" between values and plan-execution as there is between reward expectancy and behavior. Possibly the authors themselves have been trapped by the multiple meanings of the word plan. As defined their Plan consists of instructions on how to perform a certain action or process. In filling their "theoretical vacuum," however, they may be using plan in the sense of something we should *like* to do in the future.

T. A. R.

Development of the Perceptual World. By CHARLES M. SOLLEY and GARDNER MURPHY. New York, Basic Books, Inc. 1960. Pp. xiv, 353. $6.50.

In spite of its overambitious title, this book deals strictly speaking neither with development nor with the "perceptual world"; rather it attempts to sketch the outline of a theory of perceptual learning, based largely on the role of reward and punishment. As such it represents an interesting addition to the growing roster of perceptual-learning theories, to be placed alongside those of Hebb, Gibson, and the transactionalists. It is to be noted that, notwithstanding the authors' repeated professions of agreement with, if not allegiance to one or another of these rival theorists, Solley and Murphy's theory departs fairly radically from its predecessors.

The authors start from a concept of perception which is much broader than that to which many perceptual theorists would adhere. For them the process of perception is to be represented as an "act" extending over time, which can be broken up into a series of components: expectancy, attention, reception, trial-and-check, internal arousal and finally the structured percept. This schema forms the framework for Part II of the book, in which each of these stages is discussed and analyzed in turn in terms of the ways in which it may be altered through learning or experience. Since this part is only loosely connected with Part I (which presents the elements of the theory), and since it impresses this reviewer as perhaps the more valuable portion of the book, it will be considered first.

Under *expectancy* the authors consider the effect of set on the reception of information, regarded as one of narrowing alternative responses and increasing the probability of the emission of a rewarded response. It is shown, however, that expectancies require strong overlearning to become effective as determinants of perceptual reports.

In the chapter on *attending*, the "dominance principle" is emphasized: perceptual acts can be ordered along a hierarchy corresponding to the relative attention values of the different aspects of the stimulus, determined, it seems, through selective reinforcement of specific observing responses, by a process of operant conditioning. The insistence on the all-or-none character of attention appears somewhat one-sided (one wonders about the phenomenon of incidental learning, for instance), but the authors' conception of attention as not only a selective but an integrative process is an original one, worthy—as many ideas in this book—of more detailed and explicit treatment than it is given.

Skipping over the chapter on *sensory learning* (concerned with phenomena of sensitization and habituation), we turn to the stage of *trial-and-check*, which Solley and Murphy regard as the locus of much perceptual learning occurring in the absence of external reinforcement. They do not fully succeed in ridding this concept of its somewhat mysterious character, but their suggestion that it entails a process of checking the percept against schemata laid down through memory traces of previous percepts may be a helpful step forward.

The discussion in the last two chapters in this series appears to this reviewer to be too exclusively oriented around the effects of reward and punishment to do full justice to the problems involved. Thus, with respect to the first of these chapters, dealing with mechanisms of *internal scanning and feedback*, the work of Piaget on "centration" and Wapner and Werner's "sensori-tonic" research come to mind as instances of phenomena involving perceptual changes not readily fitting into Solley and Murphy's approach (though both are referred to in other contexts). The other chapter, on the *final structuring of the percept*, deals only with the problem of figure-ground articulation; a welcome feature, however, is the specification given of the conditions for the appearance of autistic effects in this type of situation.

This leads us back to Part I, where the authors present their "autism theory" of perceptual learning. Like many theories, this one appears to be in many ways hemmed in by the rather limited experimental paradigm on which it is built, based on situations which involve competition between alternative percepts, e.g., figure-ground articulations, perspectives of the Necker Cube, etc. This is reflected, for instance, in the treatment of perceptual development, devoted almost entirely to the changing role of affect and autism in the perception of the child; changes in the constancies, the illusions, the perception of shape, and other tasks where veridically is a relevant criterion, are ignored altogether.

Yet the authors repeatedly concern themselves with the relationship between autistic and veridical perception. They take the superficially not too implausible position that veridical and nonveridical learning represent products of one and the same process (cf. pp. 18, 72, 137), the latter being according to them simply more accessible to study in the laboratory. At the same time they appear to recognize the role of practice, even in the absence of eternal reinforcement, in effecting changes in perception. In order to fit this type of perceptual learning into their mold, the authors propose a parallel between the effects of practice and reinforcement on perception, but on closer analysis this seems to amount rather to a statement of their mutual interaction. One would have wished for a more thorough exploration of this intriguing notion.

If on the perceptual side the scope of the theory may thus appear too narrow, on the response side it may perhaps be criticized as too wide. Certainly no consistent attempt is made to differentiate perceptual learning from learning occurring at the response end: behaviors such as children's drawings of Santa Claus and the filling in of missing letters or words in a sentence are treated as direct manifestations of perceptual processes, and imprinting is said to be based on an animal's "innate disposition to learn to perceive something as intrinsically 'good' " (p. 128). Indeed, in their concluding chapter Solley and Murphy state that an earlier version of their formulation considered all learning as attributable to perceptual changes; now, however, they confess to a change in their views, according to which perceptual learning is to be subsumed as a species under a larger genus of acts undergoing modification. Having said this, they promptly enshroud their newly charted course in fog by suggesting (a) that the distinction between perceptual and motor acts may not be tenable, and (*b*) that the possibility of a perceptual learning *sui generis,* of an S-S type, is also to be entertained, along with the others.

In the face of such a high degree of tolerance of ambiguity, it will not come as much of a surprise to find the authors espousing with disarming frankness an eclecticism that at times reaches awesome proportions, as in their discussion of the problem of consciousness. This approach hardly seems calculated to contribute towards a clarification of the difficult question to which it is applied; indeed it tends rather to obscure basic issues, leading the authors to minimize, if not actually deny the very substantive differences between their own views on perceptual learning and those of the Gibsons, for instance.

In the opinion of this reviewer, these features of the authors' writing rob their formulation of much of its potential for generating hypotheses and stimulating research. Yet, in spite of these criticisms, the book represents an impressive effort, if only for the magnitude of the task which the authors have set themselves, and the attempt they have made, particularly in Part II, to deal with some difficult and often neglected aspects of the perceptual process. Thus, for all its shortcomings, this book deserves to be widely read on the part of psychologists concerned with problems of perception.

Clark University JOACHIM F. WOHLWILL

Men Who Manage: Fusions of Feeling and Theory in Administration. By MELVILLE DALTON. New York: John Wiley and Sons, 1959. Pp x + 318. $6.75.

This is a descriptive and an analytical account of the gap between unofficial and official action in bureaucratic industrial organization. In the words of the author, the book is "a study of compromises among key individuals and groups in rational organizations, and the human strictures on compromise." His methods are those of the non-quantitative sociologist or social anthropologist, using participant observation and living-in techniques to obtain information about generally known but relatively unrecorded subjects. Topically, the book deals with some of the most important aspects of organizational life: power distribution and struggles in the line; staff-line conflict; informal and unofficial union-management relations at the local level; the managerial career ladder; the supplemental reward structure; the connections between formal and informal action; and the designed ambiguity of the role of the manager. Few social scientists have given as much of their time and ingenuity in trying "to get close as possible to the world of managers and to interpret this world and its problems from the inside as they are seen and felt at various points and levels." It is a uniquely rich and detailed picture of the unofficial and informal internal affairs of organizations.

Men Who Manage organizes the experience, insights, and conceptualization of more than a decade of research by Dalton. Four firms with work forces from 400 to 20,000 (three factories and a department store) were the sites from which he drew his case materials. One fractionating-chemical plant of 8,000 workers and over 225 decision-making officers is described more fully than the others. This was one of the two firms in which Dalton had worked as a member of the staff and was able to note problems of recurring concern to personnel from all levels. Some of the observations of the problems regarding industrial controls, staff-line differences, and unofficial union-management relations had already been presented as progress reports in sociological journals by Dalton. The current volume, written for a more general audience of industrialists and academicians, provides a much more integrated look at the author's point of view and methodology.

The principal argument is, of course, that to understand behavior of individual managers and groups in bureaucratic structures it is necessary to look at both the formal-rational-technical and the expressive-affective-political phenomena involved. Dalton focuses on the second of these and documents in chapter after chapter the extent to which the latter accounts for departures from what might be expected if only the formal-rational side of the organization were present. There is no question but that he makes his point and well. His case materials are excellent. More importantly he goes beyond the basic data, organizing, interpreting, and formulating hypotheses for further testing. The volume is really a book of informally stated hypotheses.

There are a number of topics and concepts well worth the attention of those interested in the behavior of individuals in formal systems. The discussions of the elastic role of the assistant-to, the typology of cliques, the dichotomy of systematizers and adapters, the "two-way funnels" in informal communications systems between factions, and conflict between principle and action in organizations are just a few examples. Many of the hypotheses and concepts must be followed up in more specifically targeted research.

The methodology employed in this extended investigation of these officially uncommunicable problems is both the strength and weakness of the volume. This has been the work of a lone researcher working with no explicit hypothes at the outset, but simply trying to understand what factors account for the "schisms and ties between official and unofficial action." He obtained most of his information firsthand, but covertly, as he worked at regular staff jobs in two large plants; other data he obtained through "intimates—those who gave information and aid that, if generally known, would have jeopardized their careers." The intimates knew only of his general interest in all kinds of personnel problems from many firms; they included line superintendents, general and first line foremen, staff personnel, workers, secretaries, and husbands of secretaries. Secretaries and clerks he found were particularly helpful, being fitted both by personality and role to see certain types of events (social) and records (probably confidential), to have considerable knowledge of unofficial activities and developing policies, and to be communicative with those who show awareness of their insights and knowledge of affairs, and not treat them as intellectual menials. Dalton used his formal roles and contacts to gain access to personnel files and other records (income). As a staff member he had great freedom of movement and contact, and he learned to spot and utilize malcontents and to catch men in or near critical situations when they were less guarded in what they said. He kept work diaries and did the usual out-plant and in-plant socializing to understand better the plants and their personnel.

Dalton's brief account of his method is very helpful in the evaluation of the book and is obviously very provocative. It helps us obtain a better perspective on a basic question to which the book is indirectly addressed: how typical are these informal unofficial problems in the organization. This use of the critical incident technique on the organizational level could be predicted to identify a good deal more exceptional behavior rather than the usual within rule and policy behavior of an organizational system. At another level it raises a quite different order of question: what are—if any—the ethical standards of the social scientist in the collection of data? This latter question is much larger than can be handled in the space of this review, but it is obvious that some of Dalton's techniques for getting information make this reviewer quite uneasy and at the same time quite intrigued. It is possible that these covert procedures are the only way to study some highly taboo topics in industrial organizations. There is no question but that these or similar infiltration techniques will be required if we are to study the organization of the KKK, the Mafia, or other highly secret organizations. Methodologically, those interested in organizational research have a lot to learn—and a lot to think through ethically. Dalton's volume makes a distinct contribution in this area by resharpening old issues for us to consider in the new context of organizational research.

University of Michigan FLOYD C. MANN

Human Factors in Jet and Space Travel: A Medical-Psychological Analysis. Edited by S. B. SELLS and CHARLES A. BERRY. New York, The Ronald Press Company, 1961. Pp. xv, 386. $12.00.

The editors of this book, both of whom have contributed extensively to the field of aerospace medicine and psychology, have brought together a current account of the human implications in this new field of exploration and travel. The psychological problems facing these interdisciplinary fields receive a generous allowance of attention especially in regard to selection, fatigue, skill, systems analysis, accidents, and human engineering. The 13 contributors have provided a wide variety of ideas and new material applicable to both flight crews and passengers in jet and rocket propelled vehicles at all altitudes. Furthermore they have accomplished all of this in admirable fashion.

The book begins with some thoughtful, philosophical generalizations about manned jet and space flight and about the physician's role in this scheme. "the flight surgeon's primary responsibility is for man in actual flight." The author, Dr. Ashton Graybiel, succeeds in observing the large, integrated picture of jet flight with its human operator, and from this he extracts the role of the flight surgeon, which "is a contest between man's ingenuity on the one hand and the hostility of the environment on the other." The analysis is a good one.

The nature and elements of skilled human performance of a complex task and the deterioration of that skill (age and fatigue) are well presented. The author, Dr. William A. Wilbanks, considers several key aspects of task performance including its complex nature, its resistance to analysis and the role of perception. The usefulness of information-theory as a tool in this investigation is briefly noted. The chapter on operator-performance by Dr. W. G. Matheny, which follows this discussion, offers a real example of search for a good design of the operator's displays. The author uses three actual instrumentational projects as examples to illustrate his arguments.

Crew selection and training for jet and space flight are considered in several parts of the book. The demands placed on the human operator of new, high performance jet aircraft are determined by Dr. C. A. Berry as significantly greater than those faced by pilots of slower vehicles. He admits, however, some gaps in the accurate assessment of the tasks to be performed. With this upgraded job analysis, he suggests the importance of physical standards and physical evaluation of airmen. He emphasizes the need for physiological training of airmen, especially where oxygen supply is jeopardized. His analysis of dysbarism, the symptoms resulting from evolved gases with extreme changes in barometric pressure, constitutes an original and informative part of the chapter. Elsewhere in the book are offered current suggestions as to mental, physiological and physical requirements for space travel. Proficiency in handling high performance jet aircraft is deemed an asset, and test pilots are suggested as desirable candidates.

The chapter by Dr. S. B. Sells on the problems of group behavior of air crews in a flight vehicle is excellent. He has, in fact, applied some of the findings of small group psychology to aerospace flight. The broad subject of superior-subordinate relationships is considered. A method of analysis of real situations via group goals, group staffing, and group utilization is offered. Efficient utilization of personnel is discussed in some detail while considering the man-machine system interface.

Two subjects which receive special attention are aircraft accidents and public health measures of preventive medicine for jet aircraft. In spite of the title of Dr. Norton's chapter ("Preventive Medicine in Jet and Space Flight"), there is no real discussion of space flight. The chapter on accidents by the late Dr. H. G. Moseley, is of interest of psychologists in that he was thoroughly versed in the role of human factors in accident causation.

There are two chapters which deal with the physical properties of the upper atmosphere and the interplanetary medium. The engineering aspects of orbital and ballistic flight are discussed by A. M. Mayo including celestial mechanics, propulsion systems and structures. Dr. Hans G. Clamann provides a good, practical discussion of the physical and chemical characteristics of a number of absorbers of human waste products, and the generators of oxygen, water and nutrients. An estimate of the requirements for each occupant of a closed cabin is presented.

One of the best, as well as the most comprehensive and detailed chapters of the book is by Dr. Thrift G. Hanks. In contrast to many of the others, it contains a fair amount of quantitative data. The list of subjects discussed is a very broad one. Of particular interest is a list of common toxic agents associated with jet aircraft operation, the hazard from decompression in various sizes of aircraft cabins and quantitative estimates of oxygen requirements for aircraft, following decompression. There is a discussion of ionizing radiation hazards with jet operations, a general discussion of microwave radiation, and a quantitative estimate of the hazard of noise emanating from jet aircraft. The author concludes the chapter with a general discussion of theories of the neurophysiological structure of the central nervous system as it influences perception, motivation and response. This chapter symbolizes the number of disciplines called upon to serve aerospace medicine.

This book has a good general survey of current problems in jet and space travel. The subject matter is presented largely in descriptive rather than analytical form. This provides for simplicity and clarity of expression and a style easy to read for both the layman and the technician. There are a number of introductory chapters on subjects such as radiation and celestial mechanics which should aid the reader who is unfamiliar with aerospace medicine. No attempt was made to order and classify the material in the various chapters, which makes some repetition inevitable. The epilogue offers some philosophy about the subjects and points up the "faith" in the destiny of space exploration. He prophesies that: after one man goes into space, other men will go. There are selected bibliographies at the end of each chapter and the comprehensive subject index also will be found useful.

Harvard University ROSS A. MCFARLAND

Psychology of Literature: A Study of Alienation and Tragedy. By RALPH J. HALLMAN. New York Philosophical Library, 1960. Pp. 262. $4.75.

For readers who enjoy the process of drawing analogies, of comparing, that is, two picture-systems of mental processes, Hallman's book, *Psychology of Literature,* will be entertaining. The "psychology" he handles is of the sort that uses the structure of myth as a parallel to the more elusive aspects of thought and experience, and the "literature" includes some of the great writings, such as the *Book of Job, The Brothers Karamazov, Hamlet,* which, being of a complexity on a par with that of human behavior, offer the chance of many differing interpretations. The myth around which he groups his psychological comments is the familiar one of the death-rebirth cycle in vegetation. His, aim, as a critic, is to interpret if he can the theme of tragedy as in essence a form of the ancient myth.

In the course of the comparison—which is made in detail about Dostoyevsky's *The Possessed* and Eliot's *The Cocktail Party*—the leading idea consists in an application of the notion of "alienation." The individual, as the doctrine is expounded, is at odds with the world and with the society around him and can be brought back into conformity only by means of ritual and sacrifice. He is estranged from the simple happiness of emotional existence by his habit of thinking and of putting his thought into words. In a modern industrial society, as Fromm has been fond of pointing out, man seems to be alienated from himself by having to work under circumstances that leave him less free and joyful than he would like to be. As Hallman puts it, "perhaps the greatest alienating power of toil is to destroy ritual which alone can insure the necessary renewals and growth of man's psyche. Work makes everyday life profane and non-symbolic. . . . Labor can have therapeutic value when its purpose is for atonement rather than economic production" (p. 123).

The point being made is this, that any interpretation of man which is based directly on a myth as broad in its implication as the vegetational myth, may be full of interest but, if carried too far, may do an injustice to man's variety and to the complexity of great literature. It acts, in other words, as an alienating principle, compelling the user to fasten a rigid structure of dogma on man, who is rather fluid and non-compressible, and on literature which shares this fluidity and non-compressibility. Even the vegetational myth itself is only one fashion of comment on the rituals and picture-structures of mankind. Though it happens to be widely applied as a means for interpreting those rituals and has been used as a convenient parable of modern life, it assumes more than we yet know about the practices and interests of early communities. Hallman, for instance, makes the statement that: "It is not difficult to see that man's most dramatic and crucial experiences revolve around the disappearance in the winter and reappearance in the spring of vegetation, his chief source of food" (p. 53).

Yet, as we suppose about the rituals connected with the worship of Dionysus (whose name Hallman consistently misspells and about whose worship we know far less than we should), the maskings, dances, and mimickings were more thaumaturgic and erotic than vegetative. After all, when the wine harvest is in, how better to use the grape than for a few days of high excitement?

As a result of the very breadth of the theme underlying this book, it tends to offer comments on some great pieces of literature to which a mere student of literature would object. For example, Hallman has this to say about *Hamlet:* "The tragedy of Hamlet was that he lacked a Devil. If at any time during the course of his soul-searching soliloquies he had confronted himself with the form of the Devil, the Father's ghost would have dissolved away" (p. 151). Unfortunately for this statement, Shakespeare seems to have furnished at least one devil, his uncle, for Hamlet to reckon with, and the young man is at one point unsure as to whether the ghost he has seen may not also be the devil. In fact, with several forms of deviltry around him, the young man is hard put ot it to know what to make of things—as the best of us would be.

Whether the idea of tragedy benefits or loses from a comparison with the death-resurrection cycle is also open to question. Hallman feels that the tragic hero must welcome death in order to be free of his troubles or to expiate the guilt within him. In welcoming death as a bride, he may renew himself, may escape from feelings of inadequacy, and, by renouncing the physical life, pass beyond the limitations of the body. If this is so, then the poignancy of the world's great tragedies evaporates into air—or into theory. For the essence of tragedy is precisely as Cleopatra's servant in bare magnificence delivered it,

"The bright day is done, and we are for the dark. The consolation of tragedy is that the day of human activity, however brief its passage, is never the less justified of its brightness."

Washington University ALEXANDER M. BUCHAN

Experimental Psychology. By BURTON G. ANDREAS. New York, John Wiley and Sons, 1960. Pp. xii, 595. $6.95.

General Experimental Psychology. By LAWRENCE M. BAKER. New York, Oxford University Press, 1960. Pp. xiv, 420. $6.00.

These two texts are offered to teachers of the undergraduate experimental course.

They differ greatly in emphasis and coverage. The Baker book is the more elementary of the two and offers a skimming of certain well-worn areas of experimental psychology. The Andreas book, on the other hand, is more scholarly and detailed, and in its emphases departs radically from tradition, introducing new topics and ignoring important areas.

The book by Baker is a broad survey with chapters ranging from reaction time to motivation, including five chapters on sensation (general sensitivity, cutaneous functions, position and movement sensitivity, vision and hearing), four chapters on aspects of learning (a general chapter, one on conditioning, one on verbal learning and one on exploring, manipulation and thinking), and chapters on temporary set, association, emotion, selective aspects of perception as well as additional chapters on general methods (including psychophysics and statistics) and on conducting an original experiment. The coverage is not detailed, purposely so as indicated by the author in his introduction. The reader cannot help being struck by the absence of data and the absence of reference to specific experiments, particularly recent ones. Much of the writing might be described as 'anecdotal' with an apparent attempt to make the topics of experimental psychology relevant to the daily life of the undergraduate. For further reading, the undergraduate is frequently referred to elementary textbooks, particularly those by Munn and Morgan, as well as to experimental texts like Postman and Egan, Woodworth and Schlosberg, and Stevens' handbook. Indeed, there is more concrete experimental psychology in the better undergraduate elementary texts than is represented here.

The most distinguishing feature of the Baker book is the reproduction of photographs of the apparatus produced by the major apparatus companies for use in the undergraduate course. There are 17 such photographs. These may help the student visualize the equipment used in the laboratory course.

The book is marred by a rather careless handling of references. The name of Rubin, in reference to his figure-groundstudies, is referred to as *Rubens* in the text and index and as *Rubins* in the list of references. A reference to an article by Laidlaw and Hamilton on p. 126 has a mixed-up title and the journal cited, the *American Journal of Psychology,* is not the true one for this article, but for the article listed immediately above it in Geldard's *The Human Senses,* one by Krakauer and Dallenbach. There are also numerous omissions, misspellings of names, wrong dates and page numbers in the other references cited.

The Andreas book is divided into two parts. The first part is on general experimental methodology and it includes chapters on designing and conducting experiments, on measurement and statistics (two chapters), on psychophysical methods, on scaling, on writing research reports and on the relation between theory and research. This section is excellent and could be used in other courses besides the general experimental one. The second portion of the book deals with "selected areas of investigation." The 11 chapters range from vision and visual perception to problem-solving and social processes. The emphasis throughout is on recent research and selected experiments are treated in detail. An unusual feature of the text is its emphasis on the topic of perceptual-motor behavior. Two chapters are devoted to it and further perceptual-motor research is described in the chapters on retention and the transfer of training.

The text deliberately omits any reference to animal research on the grounds that undergraduates are often not permitted to experiment on animals. The text is also short on theory, despite a chapter on the relation between theory and research, and the topics selected are seldom treated with enough context to be meaningful. The instructor will need to supplement the research topics with background material for the student to obtain maximum benefit from them. The type of experimental psychology presented by Andreas might best be described as engineering experimental psychology.

In summary, the Baker text seems to be too slight for consideration in any serious undergraduate experimental course. The Andreas text has possibilities for those who wish a text slanted toward human engineering experimentation. Such users should be prepared to fill in the 'reason why' in lectures, where applicable, to supplement a somewhat bare presentation of research studies in the text.

George Washington University RICHARD D. WALK

Physiology of the Retina and the Visual Pathways. G. S. BRINDLEY. Edward Arnold, London, 1960. Pp. xi, 298. $7.50.

Dr. Brindley, a physiologist at Cambridge University, has written the sixth in the Monograph Series of the Physiological Society edited by Barcroft, Bayliss and Hodgkin. He takes as his theme the question "How does the human visual pathway work?" He says, in a prefatory statement, that his book is "tightly written . . . very full of information and for this reason not easy to read from end to end consecutively."

The first three chapters deal almost exclusively with electrophysiological and biochemical matters: retinal photochemistry, electrical activity of the retina, and the central visual pathways. It is impossible to summarize this material in a brief review. Numerous experiments are cited, the information is derived from a variety of species, mostly, of course, non-human, and the information, through valuable, is piecemeal, for to date no integrative picture of either human or animal visual physiology is yet available. On the whole, these sections of the book therefore resemble most strongly a comprehensive annual review in which the author has attempted to bring specific problems up to date. An example of a problem that is dealt with intensively and analytically is the origin of the electroretinogram, with emphasis on Brindley's own work and that of Tomita.

Chapter IV is an introduction to sensory experiments, which comprise the remainder of the volume. Brindley divides sensory observations into "Class A" and "Class B." Class A experiments are those in which perceived equalities or just-discriminable differences are established between different stimuli presented under identical conditions. Such experiments are acceptable for use in establishing "psycho-physical linking hypotheses," and in the "conservative" view, the only ones that can be so used. Everything else, presumably including the whole study of sensory quality, the scaling of sensory magnitudes, etc., is inadmissible in the "conservative" view, and in Brindley's own view, admissible only if it can conceivably also be tested by a "Class A" test. The author recognizes that the conservative view is clearly too restrictive for psychologists, and the present reviewers are of the opinion that the organization of the sensory material in this book would have been improved had the author been less concerned with the classification of experiments and more concerned with the phenomenal or parametric relations of the subject matter he has chosen to include. Chapter V, for instance, is entitled "Some Sensory

Experiments Involving Neither Matching Nor Threshold," and includes such heterogeous material as the perception of retinal blood-vessel shadows, additivity of brightness, the Bezold-Brücke phenomenon, the "waterfall" illusion, stabilized retinal images, and electrical phosphenes. It might have been better had the author ignored this non-A material completely, since his statement, for example, that no quantitative deductions have been made from any theory for the Bezold-Brücke effect, makes it clear that he has disregarded recent literature in these areas.

The remaining two chapters on sensory phenomena summarize experimental researches on visual thresholds and those aspects of color vision that have been explored by matching and threshold techniques. Although there is a section called "The effects of adaptation on apparent colour," it includes no mention of perceived color changes, nor are the matching data that are discussed related even inferentially to the problem of perceived color constancies. Almost without exception, qualitative color words (red, green, etc.) seem to be admissible in this chapter only (1) to specify the *stimuli,* for example, "adaptation to very bright yellow light," where the rigorous psychophysicist would insist on physical stimulus, rather than psychological response, terms; and (2), to characterize mechanisms, for example, "blue-sensitive receptors," meaning only inferred receptors that are maximally sensitive at short wavelengths.

This book will be of interest mainly to psychologists who are primarily concerned with the study of visual problems, particularly sensory mechanisms. Many of them will find the point of view narrow, but within this point of view, the "Class A" material is fairly up to date and well summarized, and the strictures of the need for clarity about assumptions and relevant evidence by theorists are apt.

New York University LEO HURVICH
DOROTHEA JAMESON

Soviet Psychology: A Symposium. Foreword by R. B. WINN. Introduction by HANS HIEBSCH with contributions by A. A. SMIRNOV, A. N. LEONTIEV, N. F. POSNANSKI, G. S. KOSTIUK, and A. L. SHNIRMAN. New York, Philosophical Library, 1961. Pp. 109. $3.75.

This symposium presents in Soviet psychologists' own words their statements on many basic issues of interest to Western psychologists. For example, Smirnov in his selection entitled "Development of Soviet Psychology" explains at length why Soviet psychologists have rejected mental testing. In brief, he argues that the prognostic use of tests belies the individual's capacity to develop his abilities and the diagnostic use of tests fails to recognize the importance of motivation. To take another example, Leontiev in the first of his two selections, "The Present Tasks of Soviet Psychology," discusses the importance of Lysenko's views to Soviet psychology. Posnanski elaborates on this in his paper, "Heredity and the Materialist Theory." Then the problem of personality is discussed by Leontiev in his second selection, "The Intellectual Development of the Child," by Kostiuk in "Problem of the Child's Personality Formation," and by Shnirman in his paper, "Investigations of Pupil Personality." These authors give no formal definition of personality, but one does get a feeling that personality is a conglomeration of traits and characteristics, such as activity, perserverance, initiative, thirst for knowledge, cooperativeness, modesty, etc.

Of equal importance to the serious student is consideration of some of the basic philosophical notions of Soviet psychology. Smirnov considers the ever present struggle between materialism and idealism in psychology. This is elaborated in more or less detail by all the contributors. The problem of consciousness is also treated by Smirnov and again by Shnirman. Indeed, the former at one place defines psychology as the study of consciousness (p. 10). Both Smirnov and Leontiev argue for the use of experiments in natural settings as opposed to the abstract experiment. Their statements may help us understand the nature of some of the contemporary "experimental" work in Soviet psychology which seems to sacrifice rigor for face validity.

The descriptive coverage of topics and the non-analytic nature of the arguments might better be understood in terms of the audience for which they were written. Unfortunately, nowhere in the book is mention made of when, where, and for what purpose these papers were delivered. Even a mildly critical reader must state that the authors have not convincingly represented their points of view. For example, the supremacy of dialectical materialism is asserted but not proved in spite of Smirnov's remarks: "Only dialectical materialism has a correct concept of the nature of experience. It proved that mental life is a special property of highly oragnized matter, which consists in representing the material world" (p. 16).

The discussion of Lysenko is another case in point. In a teleological fashion it is argued that a Lysenko approach has important implications for psychology and it is implied that the approach therefore is a more legitimate one for psychology.

The authors are generally agreed that capitalist bourgeois psychology adopts methods, *e.g.* testing, and the philosophy of idealism which support the class structure and the ruling class interests. Again the case is a polemic rather than solid evidence.

The concentration in the papers selected on personality and the Lysenko issue, the very militant style, and the fact that the latest reference cited was published in 1949 suggest that the papers are about 10 years old. A more representative collection of papers of contemporary Soviet psychology might have been selected as well as ones presenting a more sophisticated statement of the same basic views. On the other hand, since many of the issues discussed are basic, they are till of some relevance and interest.

In addition to the omissions in editing noted above, there is no statement of editorship, the indication of translator is ambiguous, and no mention is made of the status of the contributors. All these deficiencies, far from being technical details, materially detract from proper interpretation of a book of this type.

HERBERT PICK
University of Wisconsin

Training for Clinical Psychology: Proceedings of the Springfield-Mount Sinai Conferences on Intern Training in Clinical Psychology. Edited by MICHAEL H. P. FINN and FRED BROWN. New York, International Universities Press, Inc., 1960. Pp. viii, 186. $4.00.

This book is composed primarily of 12 papers read at the first Springfield-Mount Sinai Conference on Intern Training in Clinical Psychology, held in March 1955. Added to these are a paper on psychotherapy prepared specifically for the book, and a preface and "Overview" by the editors. The subtitle implies that material will be included from the second conference, which was held in March 1956, "for the purpose of delving into some of the problems raised at the first conference" (p. 4). The editors do remark on the "vigorous" discussion and occasionally attribute to the participants a point of view that is not detectable in the printed papers. Otherwise the proceedings of the second conference and the discussion from the first conference are lacking. Since the papers raise many problems and present conflicting points of view on some subjects, these omissions are regrettable, though perhaps unavoidable.

Despite these and other limitations, this book does have unquestionable value for anyone concerned with training in clinical psychology, particularly internship training. In general this reviewer agrees with the editors: "Even though the Springfield-Mount Sinai Conferences took place in 1955 and 1956, their material is as fresh and thought-provoking now as then" (p. 169). A host of practical, concrete suggestions are scattered throughout, some of which are bound to appeal and be useful to any reader. The topics cover a wide range and the participant authors represent a variety of points of view and, more importantly, bring to bear a real wealth of practical experience in clinical training programs.

The first paper, by Thomas W. Kennelly, concerns the selection and placement of psychology interns. Next, Renata A. Calabresi discusses the kind and amount of structure desirable in the training situation. Elaine F. Kinder has the topic: "Areas of Training and their Specific Implementation: Research, Therapy, Testing." The next three papers, by John W. Stafford, Gordon F. Derner, and Joseph Shoben explore various aspects of the university's role in clinical psychology training and the relationship between university and training agency. Following are three papers on problems of supervision, by Benjamin Pope, Doris Siegel (social worker), and Abraham M. Zeichner. Edward W. Slockbower's paper is titled "Problems of Communication in the Training Situation." Jules D. Holzberg discusses research training. Next is the added paper on psychotherapy training, by William Languth, a psychiatrist with considerable experience in supervising psychology interns. The final conference paper, on policies and standards for evaluating intern training programs is by Bruce V. Moore.

By and large the papers are well written and thorough and concise in their treatment. The authors discuss philosophies of training and abstract principles, but also call on their own experience to illustrate points with examples and concrete details. This reviewer, however, found Slockbower's "I"-dominated style of presentation in poor taste and his orientation toward dealing with the assigned topic clouded by a personal need to impress his audience. On most topics there is a reassuring degree of agreement. Controversy, however, is evident on the subjects of the university's role in clinical training, the relationship between university and training agency, and the schism between academic and clinical traditions.

Dr. Kinder's paper is perhaps the most provocative and deserves special mention. She is clearly angry and speaks her mind lucidly and forcefully, ending with a thorough scolding of the profession for putting "expediency before quality" (p. 48), aping other professions, failing to face up to serious limitations, and seeming to offer to students "the rewards of medicine for little effort and less cost" (p. 47). Her comments may lead to considerable soul-searching on the part of many readers.

MARY G. CLARKE

University of North Carolina

Perspectives in Personality Research. Edited by HENRY P. DAVID and J. C. BRENGELMANN. New York, Springer Publishing Company, 1960. Pp. ix, 370. $7.50.

The present volume, edited under the auspices of the International Union of Scientific Psychology, consists of revised versions of several research papers originally delivered before the 1960 International Congress of Psychology at Bonn, Germany. This international assortment of research findings and orientations is intended as a companion for the previous *Perspectives in Personality Theory,* published by Basic Books in 1957. What might have been an annoying conglomeration of unrelated research reports, however, has been skillfully edited into three major sections which reflect moderate continuities in content and appropriate contrasts in techniques. Although varying greatly in length, the papers uniformly maintain a relatively high quality and include important contributions to such topics as denial, death, suicide, and person perception.

Following a fanciful and rambling, but knowledgeable attempt by Henry Murray to place recent personality research trends in historical perspective, there appears a set of six papers grouped under the heading "Explorations in Behavior." These include discussions of denial in fantasy by Daniel Miller, the psychology of death by Alexander and Adlerstein, religion and psychopathology by Mailloux and Ancona, a cybernetic information-professing model for psychoanalytic behavior theory by Blum, and a modernized Pavlovian analysis of the development of voluntary action in children by Luria, along with a brief commentary by Tomkins which reviews the preceding five papers in relation to three current frontiers of psychological knowledge—social contexts, affective response mechanisms, and information processing.

The next section, on "Person Perception," includes discussions of the role of experienced purpose in the perception of human action by From, of movement as a cue in person perception by Tagiuri, of sources of variation in person-cognition by Taft, and of conformity as a problem in person perception by Hollander. An erudite commentary by MacLeod follows which emphasizes the phenomenological analysis of person perception, deliberately stressing the term *perception* in contrast to the previous distinctions drawn in this context by Murray between perception and apperception and by Taft between perception and cognition.

The final section, called "Resources," centers around investigative methods and includes an unfortunately brief summary by Skard, Inhelder, Noelting, Murphy, and Thomae of a symposium on longitudinal research in personality development; an investigation of suicide as a socio-psychological phenomenon by Shneidman and Farberow; and two papers on personality assessment—one on objective evaluation by Brengelmann and one on brief projective methods by David and Rabinowitz. An integrative commentary on research in personality assessment by Meili completes the section.

The final chapter, a brief note on the International Congress by Nuttin, bears a title which appropriately sounds the keynote of the entire volume—"Beyond Provincialism."

Educational Testing Service SAMUEL MESSICK

THE AMERICAN
JOURNAL OF PSYCHOLOGY

Founded in 1887 by G. STANLEY HALL

Vol. LXXV JUNE, 1962 No. 2

LENGTH OF SERIES AND THE LEARNING CURVE

NANCY C. WAUGH, Harvard University

How long does it take someone to memorize a list? How many items can he recall after he has seen this list a certain number of times? Most experiments on verbal learning answer one or the other of these questions; many of them go on to specify either the length-difficulty function or the learning curve. The length-difficulty function describes number of trials to mastery as a function of the number of items to be learned, while the learning curve relates the number of items learned to the number of times they have been exposed.[1] The present study examines the relation between these two functions for the serial learning of simple material, and shows how both may be estimated from the results of the first trial.

Little of a theoretical nature has ever been said about the length-difficulty function. The best-known theory to account for this function is Thurstone's, which states that the number of repetitions needed to learn a list of *paired associates* should increase with the square root of the number of pairs.[2] This formulation has been frequently, and sometimes improperly, applied to serial learning. What Thurstone did was to integrate the rates at which errors should drop out and successes recur in a paired-associates task. The result is the following relation:

$$T = (c/k)(se)^{1/2} \tag{1a}$$

where T is the total time required to learn a set of items, and c and k are constants related to the criterion of learning and to S respectively, while s is the total number of different correct stimulus-response associations and e is the total number

---

* Received for publication December 4, 1959. This work was carried out while the author was at the Lincoln Laboratory of the Massachusetts Institute of Technology with the joint support of the U.S. Army, Navy, and Air Force. I am indebted to B. F. Green, Jr. and J. E. Keith Smith for many of the mathematical expressions in this paper.

[1] Some writers consider total amount of time rather than total number of repetitions to be the important dependent variable, and they define the length-difficulty function accordingly. See R. S. Woodworth, *Experimental Psychology,* 1938, 20-21, and J. A. McGeoch and A. L. Irion, *The Psychology of Human Learning,* 1952, 489. The present definition conforms to Woodworth's.

[2] L. L. Thurstone, The learning function, *J. gen. Psychol.,* 3, 1930, 469-493.

of different incorrect stimulus-response associations that S can make. Thurstone assumed that to each stimulus there corresponds a unique response and that S knows the set of responses and has to learn only which response goes with which stimulus. For a list of n pairs of associates, there accordingly exist n^2 possible associations, of which n are correct and $n(n\text{-}1)$ are incorrect. Thus $(s)(e)$ becomes $n^2(n\text{-}1)$, $(se)^{\frac{1}{2}}$ becomes $n(n\text{-}1)^{\frac{1}{2}}$, and T becomes $(c/k)\ n(n\text{-}1)^{\frac{1}{2}}$. The number of repetitions required to learn the list is then

$$R = (c/k)(n - 1)^{1/2}, \qquad [1b]$$

assuming that the items are presented at a constant rate and that therefore $R = T/n$.[3]

This formulation may be properly applied to serial learning if each of n serial positions is defined as a stimulus for each of n different responses, so that se is still $n^2(n\text{-}1)$. But in a series of n items with A rather than n alternative responses possible at each position, the total number of possible associations will be nA (and not n^2), of which the number of possible correct associations will again be n and the total number of possible incorrect associations will therefore be $n(A\text{-}1)$. Thus $(s)(e)$ will equal $n^2(A\text{-}1)$. Substituting in Equations [1a] and [1b],

$$T = (c/k)n(A - 1)^{1/2}, \qquad [2a]$$

and

$$R = (c/k)(A - 1)^{1/2}. \qquad [2b]$$

Equation [2b] says that the number of repetitions, or trials to a criterion of one perfect recitation, should increase with the square root of the number of *alternative* responses, and not with the *total* number of responses, in a series (or in a set of paired associates, for that matter). It is this equation, rather than Equation [1b], which Thurstone ought to have applied to the data of Lyon. Lyon memorized lists of digits ($A = 10$) and should therefore have needed a constant number of repetitions to master each list regardless of its length.[4]

In general, Thurstone's formula does not describe the data on serial learning very well even when it is correctly applied to them. As a matter of fact, the relation between length of series and number of repetitions usually looks more like a straight line than a power function.[5] This linearity, which contradicts Thurstone's hypothesis, follows readily from some assumptions that we shall make here about the learning curve. These assumptions are quite different from Thurstone's and a good deal simpler than his. The rest of this paper will be chiefly concerned with testing them.

---

[3] Thurstone's equation was actually $R = (c/k)\ (n - a)^{\frac{1}{2}}$, where a is the span of immediate memory.

[4] Thurstone, The relation between learning time and length of task, *Psychol. Rev.*, 37, 1930, 44-53; D. O. Lyon, The relation of length of material to time taken for learning and optimum distribution of time, *J. educ. Psychol.*, 5, 1914, 1-9, 85-91, 155-163. In Lyon's experiment, trials to criterion did appear to increase with the square root of the number of items in a series. It is difficult, however, to evaluate Lyon's results, since they are based on very small samples and since his experimental design was such that practice would favor the longer lists.

[5] See, for example, the data presented in the following papers: C. N. Cofer, A comparison of logical and verbatim learning of prose passages of different lengths, this JOURNAL, 54, 1941, 1-16; V. A. Henmon, The relation between learning and retention and amount to be learned, *J. exp. Psychol.*, 2, 1917, 476-484; C. I. Hovland, Experimental studies in rote-learning theory: VII. Distribution of practice with varying lengths of list, *ibid.*, 27, 1940, 271-284; C. I. Hovland and K. H. Kurtz, Experimental studies in rote-learning theory: X. Pre-learning syllable familiarization and

Our major assumption is that a person learns the same number of new items every time he recites a list. Every repetition of the list simply adds a constant number of items to the ones he has already learned. We assume further that he does not forget an item once he has learned it. His learning curve should therefore be linear. Thus, for instance, if he picks up 10 new items on every trial and he is trying to learn a list of 100 items, he should remember a total of 20 items after the second trial, and 50 after the fifth, and he should master the list on the tenth trial. Assume now that the number of items he learns on each trial is the same for all lengths of list. Then it should take him two trials to learn a list of 20 items and five to learn a list of 50. The length-difficulty function should thereby be linear.

Of course, the exact number of new items that our *S* picks up will vary from trial to trial. Sometimes he may pick up 20 new items and sometimes, none. His increment of 10 items is really an average increment. If we gave this person many trials with many lists and counted the number of new items that he learned on each trial, we would obtain a frequency-distribution of increments that ran from 0 up to k items, where k was the greatest number that he ever picked up. We could then state the probability that he would learn exactly 0,1,2, . . . , or k new items on any given trial. If the number of items he learned on one trial did not influence the number he learned on the next trial, we could also state the probability that he would recall a total of 0,1,2, . . . , or tk items, new plus old, on Trial t. Let us see how we would do this.

Suppose that our *S* learns 0,1,2, . . . , or k new items on the first trial with a probability of $P_1(0)$, $P_1(1)$, $P_1(2)$, . . ., or $P_1(k)$ respectively, where the subscript refers to the number of the trial and the quantity in parentheses, to the number of items. These probabilities, we assume, are constant from trial to trial, so that $P_1(i) = P_2(i) = P_t(i)$, for i ranging from 0 to k. Assume now that the number of items this person picks up on any trial is independent of the number he picked up on any previous trial, and that once he learns an item he does not forget it. Then the probability that he will recall exactly j items on the second trial will be

---

the length-difficulty relationship, *ibid.*, 44, 1952, 31-39; C. E. Noble, The length-difficulty relationship in compound trial-and-error learning, *ibid.*, 54, 1957, 246-252; E. S. Robinson and W. T. Heron, Results of variations in length of memorized material, *ibid.*, 5, 1922, 428-448; T. C. Scott and L. L. Henninger, The relation between length and difficulty in motor learning; a comparison with verbal learning, *ibid.*, 16, 1933, 657-678.

$P_1(0)P_2(j) + P_1(1)P_2(j\text{-}1) + \dots + P_1(i)P_2(j\text{-}i) + \dots + P_1(j)P_2(0)$. If, for instance, he always learns just either 0 or k new items on any trial, including the first, with probabilities $P(0)$ and $P(k)$, then we would predict the following probabilities on the second trial: no items recalled, $P(0)P(0)$, or none on either trial; k items recalled, $P(0)P(k) + P(k)P(0)$, or k on one trial and none on the other; and $2k$ items recalled, $P(k)P(k)$, or k on both trials. On the third trial, the probabilities would be: no items, $[p(0)]^3$; k items, $3\,[P(0)]^2\,P(k)$; $2k$ items, $3\,P(0)\,[P(k)]^2$; and $3k$ items, $[P(k)]^3$. The range, the mean, and the variance of the distribution we observed on the first trial are all increased by a factor of t on the tth trial. This rule applies not only to a binomial distribution, like the one we have just considered, but to any kind of distribution.

To put this more formally, we may define an arbitrary term, a^i, where i is a random variable that ranges from 0 to k and a is an indifferent constant. We then form the expression $P(0)a^0 + P(1)a^1 + P(2)a^2 + \dots + P(k)a^k$. Let us call $P(i)$ the probability of learning i new items on any trial. Then the probability of recalling a total of j items, new plus old, on trial t — let us call it $P_t(j)$ — will be the coefficient of the term a^j in the multinomial expansion $[P(0)a^0 + P(1)a^1 + P(2)a^2 + \dots + P(k)a^k]^t$, where a^i now ranges from a^0 to a^{tk}. The distribution of i that we get when we perform this expansion is called the t-fold convolution of the original distribution, for which the exponent t was equal to 1.[6] For our present purposes, the important thing to realize is that when we convolute a distribution t times, we increase not only its range but also its mean and its variance each by a factor of t.

The basic hypothesis, then, is that the probabilities that a person will learn 0,1,2, . . . , or k new items from a list are constant from trial to trial and combine independently on successive trials. These items, once they are learned, are recalled on all later trials. The number of new items learned on any trial is independent of the number of trials and is independent of the number of items already learned. To test the hypothesis we can simply have people attempt to learn a long list of items over several trials, and then examine the distribution of the total number of items that they recalled after each trial. If the mean and variance of this distribution for Trial t are each t times the means and variance for Trial 1, then the hypothesis is correct. Experiment I was such a test of this hypothesis.

EXPERIMENT I

Method. S's task was to memorize lists of 64 octal digits. These digits, *0* through *7*, were arranged in quasi-random sequences in which each digit appeared a total of eight times, but never more than twice in succession. A given list was exposed for six consecutive trials in each of 10 experimental sessions, with a different list for each session. The digits were shown through the window of an electric digit-display device, at a rate of 1 per sec., with an exposure-duration of 0.9 sec. They appeared as black numerals about 1½ in. high and ¾ in. wide on a white background. The end of a series was signalled by the appearance of a black *9* on a red background. *S*, who sat at a comfortable viewing distance from the display-device, initiated each series by pressing a switch, after spending as much time as he wished in recalling the previous series. He was instructed to write the digits down in the appropriate numbered blanks on a response-sheet in order from first to last, and he was told to fill in all of the blanks and to guess as much as he had to in order to do so. He was not allowed to refer to a response-sheet after he had filled it in, nor was he ever told which items he had recalled correctly.

---

[6] For a discussion of convolutions, see W. Feller, *An Introduction to Probability Theory and Its Applications*, 1950, 214-216, 218.

The *S*s were four women and one man. All were college graduates and had had previous experience in learning lists of digits. Each learned the series in a different random order. There was usually a day's interval between experimental sessions.

Scoring. The data were scored in terms of overlapping pairs of items recalled correctly, rather than of single items recalled correctly, in order to minimize the effects of guessing. The highest score possible was therefore 63, and the total score expected on the basis of independent random guesses on adjacent items was accordingly 63/64 rather than 8. A *score* thus refers to the number of pairs of items correctly recalled on a given trial, where an item was considered to have been recalled correctly only if it was the correct digit in the correct serial position. We shall denote by m_t and σ_t^2 respectively the mean and variance of the distribution of the scores observed on Trial t.

The point of this experiment was to determine whether m_t and σ_t^2 increase linearly with number of trials. The obvious procedure was to plot these quantities as a function of t and determine whether either set deviated significantly from the straight line through the origin that best fitted it. This procedure was not feasible for two reasons. First, inspection of the data revealed stable differences between *S*s in the number of items learned on each trial. The total distribution of scores was really composed of five individual distributions, and its variance was therefore equal to the mean of the individual variances plus the variance of the individual means. Thus if the individual means and variances increased linearly with t, the group's variance would increase as a function of t^2 as well as of t.[7] Secondly, two of the five *S*s mastered most of the lists on the fifth and sixth trials respectively. Either to include or omit their scores for these trials would have spuriously reduced m_t and σ_t^2. The alternative was to omit all of their scores.

The data were therefore transformed so as to eliminate the variance of the means from σ_t^2 and so as to permit the use of the scores obtained on precriterial

---

[7] Thus

$$V_1 = \bar{V}_1 + \left[N \sum^{N} m_i^2 - \left(\sum m_i\right)^2\right](1/N) \text{ and}$$

$$V_t = t\bar{V}_1 + \left[N \sum^{N} (tm_i)^2 - \left(\sum^{N} tm_i\right)^2\right](1/N),$$

where V_1 is the variance of the composite distribution, $\bar{V}_1$ is the mean of the variances of N individual distributions, and m_i is the mean of an individual distribution, all observed on Trial 1, while V_t is the variance of the composite distribution predicted by the present hypothesis for Trial t.

trials by the two *S*s who mastered the lists. The scores were, in effect, normalized and were then added to a fitted (rather than an observed) mean for the group. This step changed the data in no way essential to the present inquiry. It was taken partly in order to make maximal use of the data and partly in order to make them more readily intelligible.

Specifically, this procedure was as follows. First, lines through the origin were fitted by a method of least squares to (a) the means and (b) the variances of each individual's set of scores on successive trials. Thus the mean and the variance of his scores for the first trial were estimated from the means and variances observed on all of the trials. These individual estimates of m_1 and σ_1^2 were then averaged across individuals and multiplied by t to give a fitted mean and variance, $t\hat{m}_1$ and $t\hat{\sigma}_1^2$, for the group on Trial t. (Multiplying the estimated m_1 for each individual by t, on the other hand, gave F_i, a fitted mean score for the individual on Trial t.) For each individual, then, as well as for the group, there was both an observed and a fitted mean and variance for the distribution of scores on each trial. The individual's fitted score was subtracted from each of his observed scores for a given trial to give a difference score. This was added to the fitted group mean for each trial to give a set of adjusted scores. Thus,

$$A_i = O_i - F_i + F_g,$$

where, for a given number of trials, A_i is an adjusted individual score; O_i is an observed individual score, F_i is the fitted individual score, and F_g, which is equal to $t\hat{m}_1$, is the fitted group score.

Results. The cumulative frequency-distributions of these adjusted individual scores, A_i, are ploted in Fig. 1 on linear by normal-probability coordinates. The ordinate denotes the probability that at least A_i pairs of items were recalled correctly. The parameter is number of trials. The solid lines in this figure represent the distributions predicted by the present hypothesis. For each distribution, the fitted mean on Trial t is equal to t times the fitted mean for the group on the first trial, and the fitted variance on Trial t is equal to t times the fitted variance for the group on the first trial.

The distributions for the first four trials represent the scores of all five *S*s, while the distributions for Trials 5 and 6 represent the scores of four and three *S*s, respectively, since the expected means for one *S* on trial five and for two *S*s on trial six were above 63 (the maximal score) and therefore their adjusted scores for those trials would have been spuriously low. The fitted means and variances for each of these *S*s were based on their scores for the first four and five trials respectively.

It is evident that the observed and the estimated distributions are in excellent agreement. In no case do an observed and an estimated distribution differ significantly from each other, according to the Kolmogorov-Smirnov test for the difference between two cumulative frequency-distributions.[8] (Actually, this is a more stringent test than the hypothesis requires: it tests all of the moments of the distributions rather than just their means and variances, which are all that the hypothesis predicts. It so happened that the observed distributions were normal and hence completely defined by these two parameters.)

The present hypothesis states also that the mean number of items recalled for the first time on any given trial is independent of the number recalled for the first time on any other trial. This assumption was tested by the following procedure. First, every *S*s score on Trial *t* with a given list was subtracted from his score on Trial $t + 1$ with that list, either up to Trial 6 or up to the last trial before a perfect score was attained. From each of these differences was then subtracted the individual's fitted increment per trial (the slope of the least-squares line through the origin describing his mean score per trial). Each of these adjusted increments observed on

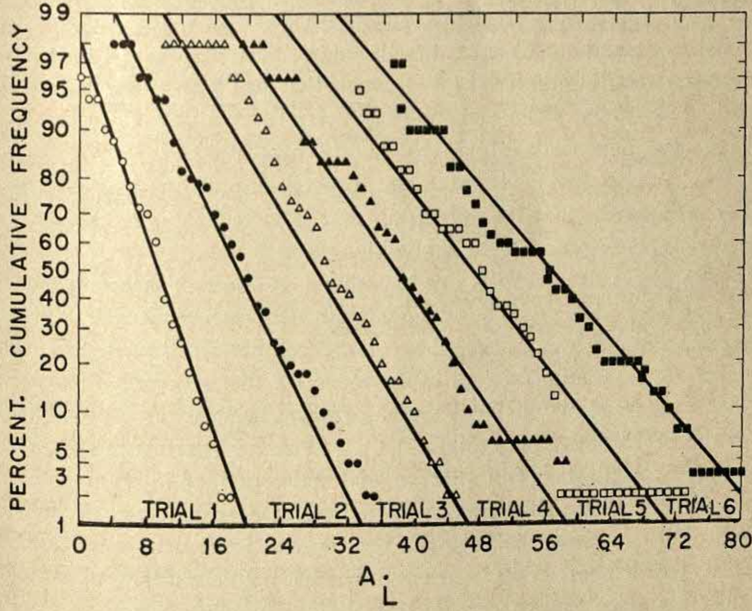


FIG. 1. CUMULATIVE FREQUENCY-DISTRIBUTIONS OF ADJUSTED SCORES FOR LISTS OF 64 DIGITS

The parameter is number of trials.

---

[8] L. A. Goodman, Kolmogorov-Smirnov tests for psychological research, *Psychol. Bull.*, 51, 1954, 160-168; N. Smirnov, Table for estimating the goodness of fit of empirical distributions, *Ann. math. Statist.*, 19, 1948, 279-281.

Trial t with a given list was then paired with the increment observed on Trial $t + 1$ with that list. The product-moment correlation coefficient computed for these pairs ($n = 216$) was -0.06, which does not differ significantly from zero. (A correlation coefficient between pairs of *unadjusted* increments would probably be high and positive, since the mean increment was different for each individual.)

Does the difference between the scores on Trial t and on Trial $t + 1$ really represent pairs of items recalled for the first time on Trial $t + 1$? It does if all of the pairs recalled correctly on Trial t are recalled correctly on Trial $t + 1$ and on subsequent trials. In the present case, the probability of correctly recalling on Trial $t + 1$ a pair that had been correctly recalled on Trial t was 0.95. (The probability of correctly recalling on Trial $t + 2$, and on subsequent trials, a pair that had been recalled on Trial t was not calculated.) The close fit between the theoretical and the observed values in Fig. 1 suggests that the forgetting that occurred 5% of the time between trials may be considered negligible.

In sum, these data support our hypothesis. A constant mean number of items was recalled correctly for the first time after each exposure of a series. The exact number was independent of the number recalled on the previous trial. The variance, as well as the mean, of the frequency-distribution of scores observed on the Trial 1 was increased by a factor of t on Trial t. Thus the probabilities of recalling $0, 1, \ldots$, and k items correctly for the first time were constant from trial to trial and combined independently on successive trials. We shall now see whether these probabilities depend on length of list; and we shall also see how we can predict the mean number of items recalled correctly as mastery of the list becomes increasingly likely, so that this mean cannot increase linearly with number of trials but must instead approach the total number of items in the list.

EXPERIMENT II

When we have a group of people recite a list over and over again, we may construct a learning curve that shows the mean number of items recalled by the group on each trial. The present hypothesis says that this function should be a straight line through the origin, but each individual, we suppose, will sooner or later recall the entire list. Whenever someone does reach this limit, he obviously cannot learn any new items on the next trial, and the rate at which the group is learning must thereupon decrease. With a relatively short list, this will happen on an early trial, while with a relatively long list, it will not happen for some time—and only if the list being learned were infinitely long could the learning curve remain linear indefinitely. With any finite length of list, it can be linear only up to a certain point, beyond which it must level off as more individuals recall all of the items. Our hypothesis must therefore be expanded to cover this fact.

Let us say that the frequency-distribution of scores obtained, by the group on the first trial ranges from j to k and has a mean of m_1. Our hypothesis states that the distribution on Trial t will range from tj to tk and will have a mean of tm_1. At some point, however, as we have said, tk will exceed r, the total number of items in the list, and an individual who would have recalled more than r items on that trial will now recall just r items. Thus the frequency-distribution of scores from that trial on will be obstructed, so to speak, by this ceiling on the maximal number of items that can be recalled. The distribution from tj to r will still be what it would have been were the list infinitely long, while the frequency at r will be the sum of the frequencies that would have occurred at all values equal to or greater than r. Such a distribution is not censored at r, for the frequencies that in that case would have occurred at unspecified values greater than r now occur at r exactly;[9] nor is it truncated, for in that case these frequencies would not appear in the sample at all.[10] This kind of distribution, two forms of which are illustrated in Fig. 2, will therefore be called an obstructed distribution. A distribution may, of course, be obstructed either below or above its mean; the resulting mean will accordingly be greater or less than the mean of the unobstructed distribution, since the frequencies that would have occurred below or above r respectively are now at r.

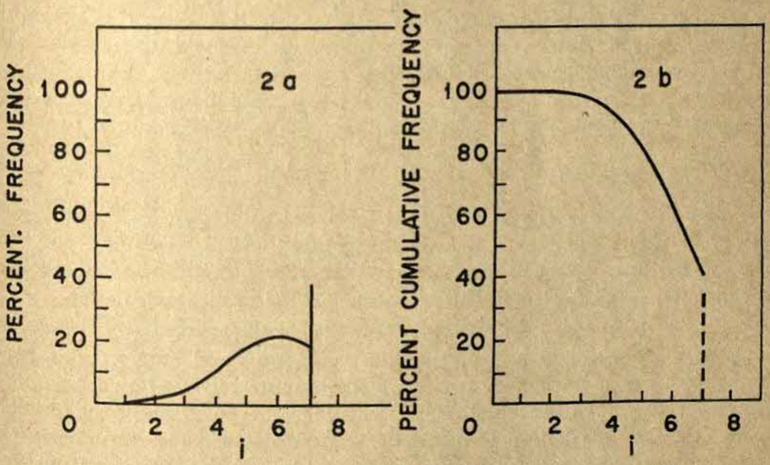


FIG. 2. AN HYPOTHETICAL DISTRIBUTION OBSTRUCTED AT $i = 7$
2a = density form; 2b = cumulative form.

---

[9] A. M. Mood, *Introduction to the Theory of Statistics*, 1950, 243.

[10] A. C. Cohen, Jr., Estimation of the Poisson parameter from truncated samples and from censored samples, *J. Amer. statis. Ass.*, 49, 1954, 158-168.

The problem is therefore to predict the mean of an obstructed distribution of scores on Trial t, on the assumption that this mean would be a linear function of t if the list being learned were infinitely long. In order to do this, we first define the following terms: M is the mean of a normal distribution; σ^2 is its variance; r is any value on the abscissa; z, its normal deviate, is $(r\text{-}M)/\sigma$; $f(z)$ is the ordinate at z; and $F(z)$ is the area from $-\infty$ to z. It can be shown that the mean, m, of a normal distribution obstructed at r is

$$m = r + (M - r)[F(z)] - \sigma[f(z)]. \qquad [3a]$$

Let us again say that r is the total number of items in a list. We assume that several persons are learning this list, and that each will recite the list for an indefinite number of trials. The present hypothesis states that with an infinitely long list M_t, the mean of the distribution of scores on Trial t, will be tm_1, and that $t\sigma_1^2$ will be the variance of this distribution, where m_1 and σ_1^2 are the mean and variance of the distribution observed on the first trial. Turning now to a list of finite length, we may define z in Equation [3a] as $(r - tm_1)/(t^{\frac{1}{2}}\sigma_1)$. Substituting in Equation [3a], we find that the mean number of items, m_t, that will be recalled on trial t is then

$$m_t = r + (tm_1 - r)[F(z)] - (t^{\frac{1}{2}}\sigma_1)[f(z)]. \qquad [3b]$$

It is clear that when r is very large relative to tm_1 and $(t^{\frac{1}{2}}\sigma_1)$, as it would be on early trials with a long list, z is large and positive, and thus $f(z)$ is 0 and $F(z)$ is 1, and m_t therefore equals tm_1. Later on —when, say, $tm_1 = r$— z is 0, and m_t is accordingly $r - 0.3989(t^{\frac{1}{2}}\sigma_1)$. Finally, when r is very small relative to tm_1 and $(t^{\frac{1}{2}}\sigma_1)$, z is large and negative, $f(z)$ and $F(z)$ both approach 0, and m_t approaches r. The distribution that we eventually observe when everyone has mastered the list is a point distribution at r. Thus the learning curve, m_t as a function of number of trials, will be effectively linear up to the point where someone first masters the list. It will then be negatively accelerated as more and more people do so and m_t thus falls increasingly short of tm_1—as it must unless the list is infinitely long or unless tm_1 is the mean of a point distribution.

We have assumed here that m_t is a mean estimated across several individuals and one list. It could, of course, just as well be a mean estimated accross one individual and several lists or across several individuals and several lists. We have also assumed that m_1 is the mean of a normal distribution. Actually, according to the present hypothesis, Equation [3b] should hold approximately regardless of the shape of the distribution observed on the first trial. The approximation will be more precise the larger the value of t, since successive convolutions of any distribution will, under fairly mild restrictions, ultimately result in a normal distribution. If the original distribution is normal, then Equation [3b] should hold for any value of t.

Experiment II was performed to determine: (1) whether Equation [3b] predicts the learning curves for relatively short lists; and (2) whether m_1 and σ_1^2 vary with the length of the list being learned.

Method. The general experimental procedure was the same as in Experiment I. The material to be learned consisted of 9 series each of 16, 24, and 32 octal digits, a total of 27 series. Each digit appeared an equal number of times in each series, and no digit ever appeared more than twice in succession. One list of each length was exposed six consecutive times in the course of an experimental session, and a different set of three lists was thus learned in each of nine sessions. The order of list-lengths within a set was either 16, 24, and 32 items, or 24, 32, and 16 items, or 32, 16, and 24 items. Each successive block of three sessions included three sets, chosen and ordered at random, which represented each of these three patterns of list-lengths. The particular sets and the particular order in which they appeared were determined independently for each *S*.

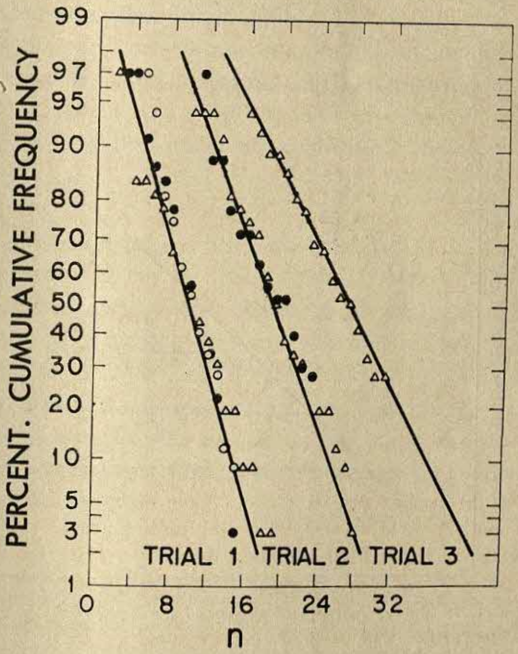


FIG. 3. CUMULATIVE FREQUENCY-DISTRIBUTIONS OF SCORES FOR LISTS OF DIFFERENT LENGTHS

The lines for Trials 1 and 2 are defined by the mean of the means and variances observed (1) for lists of 16 items (open circles), 24 items (filled circles), and 32 items (triangles), and (2) for lists of 24 and 32 items, respectively. The line for Trial 3 was fitted by hand.

The *S*s were three women and one man, all college graduates, three of whom served in Experiment I. The first set of lists learned by each *S* was subsequently defined as a practice set and discarded. The data from this experiment thus consist of eight lists of each length learned by each of four *S*s. Of these eight lists, at least seven were identical for any given pair of *S*s.

Results. Fig. 3 shows the cumulative frequency-distributions, plotted on linear by normal-probability coördinates, of scores obtained by the group on the first trial with the lists of 16, 24, and 32 items; on the second trial with the list of 24 and 32 items; and on the third trial with the list of 32 items. This missing distributions are so obstructed that there is probably nothing to be learned from looking at them. These five distributions appear to be normal, for all practical purposes, and the means and the variances for the 16-, 24-, and 32-item lists after Trial 1 and for the 24- and 32-item lists after Trial 2 do not differ significantly from each other, according to the Kolmogorov-Smirnov two-sample test.

The three lower curves in Fig. 4 represent m_t, the mean number of overlapping pairs of digits correctly recalled, as a function of t, the number of trials, for the lists of 16, 24, and 32 items. The symbols are the observed values of m_t and the lines represent the values predicted by Equation [3b], with r set equal to 15, 23, or 31. The expected values of m_1 and σ_1 are the means of m_1 and σ_1 observed for all three lists. It is clear that the observed and the expected values of m_t for each length of list are in good agreement over the course of the six trials. We may thus conclude that m_1 and σ_1 are indeed independent of list-length and that the mean number of items correctly recalled increases with number of trials in the way that the present hypothesis demands.

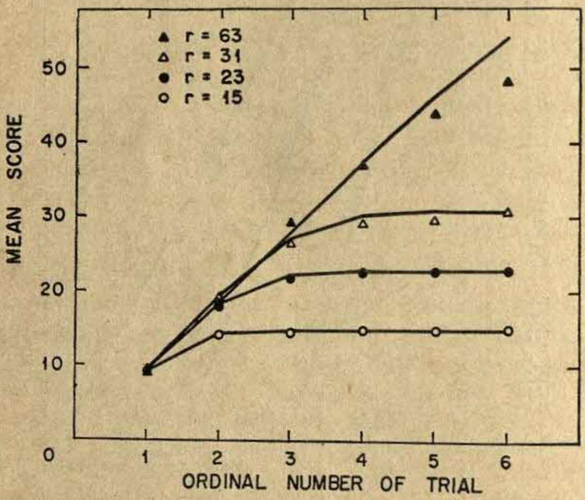


FIG. 4. MEAN NUMBER OF PAIRS OF ITEMS RECALLED AS A FUNCTION OF NUMBER OF TRIALS

The parameter is the total number of pairs in a list.

Fig. 4 shows the mean scores expected and observed with the 64-item lists in Experiment I. The expected function is based on tm_1 and $t\sigma_1$ estimated for these lists as described previously. In Fig. 4, the observed values of m for the 64-item list and, to a lesser extent, for the 32-item list, tend to lie below the expected values, although never significantly so. Note, however, that Equation [3b] assumes that the variance of the frequency-distribution of correct responses will increase linearly with number of trials. But the variance of a set of pooled individual distributions with different means will increase also with the square of each convolution of the composite distribution, a point that was discussed in connection with the transformation of the scores in Experiment I. Equation [3b] should therefore overestimate m to the extent that σ_t^2, the variance of the distribution on Trial t, exceeds $t\sigma_1^2$, which would be the variance of this distribution if all of the individual means were identical.

In this experiment, then, the mean number of items recalled from a list of r items could be described as the mean of a normal distribution obstructed at r. As in Experiment I, this mean increased linearly with number of trials up to the point where the probability of recalling an entire list became appreciable. The negative acceleration of the learning curve beyond this point varied directly with list-length. The mean and variance of the distribution of scores observed on the first trial, on the other hand, were independent of list-length.

DISCUSSION

(1) Both of these experiments have supported the present hypothesis: up to the point where *S* could master an entire list, each exposure of the list added a constant number of new items to those that he could already recall. And up to this point, the total number of items, new plus old, that he could recall increased linearly with number of trials. The negative acceleration of the learning curve beyond this point was shown to be an artifact of list-length.

So far we have been dealing with the learning curve. What about the length-difficulty function? If the learning curve is independent of list-length and is linear, as it was in these experiments, then the length-difficulty function must also be linear. Let us develop this point formally. Assume first that we have a normal frequency-distribution of correctly recalled items on each trial. Let us again, as in Equation [3a], define z, a normal deviate, as $(r - tm_1)/(t^{\frac{1}{2}}\sigma_1)$, where m_1 and σ_1 are the mean and standard deviation of the distribution of scores on the first trial and $F(z)$

is the area under the normal curve from $-\infty$ to z. The present hypothesis says that the probability that at least r items will be recalled on Trial t is

$$P_t(r) = 1 - F(z). \qquad [4]$$

In order now to find $t_{0.50}$, the median number of trials to a criterion of one perfect recitation of a list, we let r denote the total number of items in a list and we set $P(r)$ equal to 0.50. Then

$$F[(r - m_1 t_{0.50})/\sigma_1(t_{0.50})^{\frac{1}{2}}] = 0.50;$$

$$[r - m_1 t_{0.50}]/[\sigma_1(t_{0.50})^{\frac{1}{2}}] = 0;$$

and

$$t_{0.50} = r/m_1.$$

The variable $t_{0.50}$ is the number of that trial on which the probability is 0.50 that all r items will be recalled correctly for the first time; the memory span could be said to consists or r items on this trial. If m_1 is independent of list-length, $t_{0.50}$ is then simply the number of items in a list divided by the mean number recalled on the first trial. The relation between $t_{0.50}$ and list-length will depart from linearity, of course, to the extent that m_1 depends on list-length.

The mean number of trials to criterion is

$$t_\mu = \sum_{t=1}^{\infty} F[(r - m_1 t)/t^{\frac{1}{2}}\sigma_1], \qquad [5]$$

where all of the terms are defined as above. This quantity is clearly not independent of σ_1, and it is not obvious that μ should increase linearly with r. Calculations have shown, however, that for r ranging from 5 m to 20 m, $t_\mu - t_{0.50}$ will assume a very nearly constant value that depends just on σ_1/m_1. This means that the mean number of trials to criterion will equal the median number of trials to criterion, r/m_1, plus an increment that depends on the variance and mean of the frequency-distribution of the number of items recalled on the first trial. To the extent that this increment is independent of list-length, the mean number of trials to criterion will be a linear function of list-length. According to the present hypothesis, then, both the learning curve and the length-difficulty function are contained in the outcome of the first trial. Once σ_1 and m_1 are thereby known, the outcomes of later trials are, in a sense, redundant.

(2) The stochastic models that Miller and McGill and that Bush and Mosteller have proposed to describe verbal learning assume that, on each

trial, S learns some proportion of the items that are left to be learned, rather than a constant number of them.[11] According to these models, which refer to free recall, the probability that an item will be recalled increases gradually with the number of times that it has been recalled previously. The present hypothesis, which refers to serial recall, states that the probability that an item will be recalled jumps to unity after it has been recalled once. If an item has not been recalled by Trial t, on the other hand, the probability that it will be recalled for the first time on Trial $t + 1$ is $m_k/(r\text{-}d)$ for $r\text{-}d \geq k$, where m_k is the mean of a distribution from 0 to k of the number of items recalled for the first time on any trial, r is the number of items in the list, d is the number learned by Trial t, and $r\text{-}d$ is therefore the number still to be learned. The mean value of d will be tm_k as long as $r\text{-}d \geq k$. If, on the other hand, $r\text{-}d = j$, where $0 < j < k$, the probability that an item will be recalled for the first time on Trial $t + 1$ is $m_j/(r\text{-}d)$ for $r\text{-}d < k$ where m_j is the mean of the same distribution which was referred to above and which is now obstructed at j. The overall probability that an item will be recalled for the first time on any trial remains constant at m_k/r, as long as $r\text{-}d \geq k$.

The results of these experiments thus argue that repetition as such does not increase the likelihood that an item will be recalled: an item is likelier to be recalled for the first time on Trial $t + 1$ than on Trial t simply because the size of the sample of unlearned items is smaller on Trial $t + 1$ than on Trial t. It is therefore as though an item that had not been recalled on Trial t were being seen for the first time on Trial $t + 1$. This inference is consistent with Rock's finding that when new items replace unlearned ones after each trial with a list, the number of trials that it takes to learn the list is the same as when they are not.[12]

It seems unnecessary, in view of these data, to assume that memorization depends on the growth of habit strength or the decay of intraserial interference with repeated exposures of a list. Instead, it would be simpler to suppose that a person can attend to only a limited number of items within the time that it takes to present a list, and that on each successive exposure of this list he is able to attend to a new group of items. When recall is serial, the size of this group appears to be constant with different lengths of list and after different numbers of trials.

---

[11] G. A. Miller and W. J. McGill, A statistical description of verbal learning, *Psychometrika*, 17, 1952, 369-396; R. R. Bush and Frederick Mosteller, *Stochastic Models for Learning*, 1955, Ch. 10.

[12] Irvin Rock, The role of repetition in associative learning, this JOURNAL, 70, 1957, 186-193. Cf. also L. L. Clark, T. G. Lansford, and K. M. Dallenbach, Repetition and associative learning, this JOURNAL, 73, 1960, 22-40.

SUMMARY

The present study deals with the serial recall of simple material: the *S*s recited lists of digits of various lengths after each of six exposures of a list. The probabilities that a person would recall, 0, 1, or k items correctly for the first time were found to be independent of list-length and of number of trials, and they combined independently on successive trials with the same list. An item once recalled was almost always recalled on subsequent trials. The total number of items recalled, new plus old, was thus a linear function of number of trials, up to the point where an entire list could be recalled. The negative acceleration of the learning curve beyond this point was shown to be a statistical artifact. It was shown also that under these circumstances the length-difficulty function, and number of trials to mastery as a function of length of list, must likewise be linear.

STROBOSCOPIC MOVEMENT BASED ON CHANGE OF PHENOMENAL RATHER THAN RETINAL LOCATION

By IRVIN ROCK, Yeshiva University, and SHELDON EBENHOLTZ,
New School for Social Research

Although stroboscopic movement has been the subject of countless investigations over a period of many years, it remains to this day an unexplained phenomenon. For the most part, work on this problem during the last several decades has been directed at adducing evidence for or against two types of theories. The Gestaltists have hypothesized that the experience of movement is based on some central physiological interaction which takes place in the cortex between the loci of excitation yielded by the two light-sources.[1] Support for this position has been seen in the finding that the effect is superior or occurs more readily within one hemisphere than between the two,[2] in the finding that satiation of the cortical region between the loci of excitation will subsequently effect *phi,*[3] and conversely, in the finding that after-effects of stroboscopic movement can be predicted for objects which stimulate that cortical region or 'field.'[4]

Others have sought to show that the interaction is either retinal[5] or sub-cortical[6] and support for this view has been seen in the fact that the effect is more readily obtained when both sources stimulate the same eye than when each is located in a different eye,[7] and in the finding that decorticated guinea pigs apparently achieve stroboscopic movement.[8] Both

---

* Received for publication November 28, 1960. This study was supported by Research Grant M-2082, National Institute of Mental Health, Public Health Service.

[1] Max Wertheimer, Experimentelle Studien über das Sehen von Bewegung, *Z. Psychol.,* 61, 1912, 161-278.

[2] A. Gengerelli, Apparent movement in relation to homonymous and heteronymous stimulation of the cerebral hemispheres, *J. exp. Psychol.,* 38, 1948, 592-599.

[3] B. H. Deatherage and M. E. Bitterman, The effect of satiation on stroboscopic movement, this JOURNAL, 65, 1952, 108-109.

[4] N. H. Livson, After-effects of prolonged inspection of apparent movement, this JOURNAL, 66, 1953, 365-376.

[5] C. T. Morgan, *Physiological Psychology,* 1943, 203-206; Koiti Motokawa, Retinal traces and visual perception of movement, *J. exp. Psychol.,* 45, 1953, 369-377; The physiological mechanism of apparent movement, *idem,* 378-386.

[6] K. U. Smith, The neural centers concerned in the mediation of apparent movement vision, *J. exp. Psychol.,* 26, 1940, 443-466; K. U. Smith and W. E. Kappauf, A neurological study of apparent movement, *J. gen. Psychol.,* 23, 1940, 315-327.

[7] C. H. Ammons and Joseph Weitz, Central and peripheral factors in the phi phenomenon, *J. exp. Psychol.,* 42, 1951, 327-332.

[8] Smith, *op. cit.,* 1940, 443-466; Smith and Kappauf, *op. cit.,* 315-327.

theories share the common assumption that the effect depends upon a neural interaction between the excitations and that it is such interaction which creates the experience of movement.

The usual statement of the minimal conditions required to produce stroboscopic motion is that disparate retinal points must be exposed to alternate stimulation. Given the proper rate of such stimulation, the experience of motion will occur. While there can be doubt that the above statement represents a sufficient condition, one can by no means conclude on the same basis that it is also a necessary condition. The sources of stimulation—corresponding to the disparate retinal points—are typically localized phenomenally at distinct places in space. There are, therefore, two possible ways of stating the necessary conditions for movement, one in terms of anatomical locus of stimulation and the other in terms of experienced location of the sources of stimulation. It is to be noted that in the traditional statement of the minimal conditions required to produce stroboscopic motion these two possible alternatives are not distinguished, undoubtedly because change of retinal location is generally a concomitant of change of phenomenal location.

The possibility which was explored in the following experiments was that stroboscopic motion is primarily a function of change in *phenomenal location* of the source of stimulation and not at all a function of change in the anatomical locus of stimulation (neural interaction theories, as noted above, are based upon the assumption of change of anatomical locus).[9]

EXPERIMENT I

Part 1. The purpose of this experiment was to create conditions such that a single anatomical locus of stimulation be experienced at two different spatial locations. This was accomplished by having *O* so move his eye back and forth as to view each stimulus foveally. Thus, while *O* localized the distal object on his left and again on his right, the locus of retinal (and cortical) stimulation remained identical in both cases.

Procedure. O's head was held rigidly in a fixed position by a head-rest. He viewed the scene monocularly (generally with the right eye) and alternately through one of two artificial pupils. The artificial pupils were about 0.014-in. in diameter, about ⅝-in. apart, punched in a circular piece of thin black paper surrounded by a shield which prevented *O* from seeing anything except through these openings. *O* could see through only one aperture at a time, in that light stimulated his retina only through the aperture his eye was facing at any moment.

---

[9] Of course, phenomenal location depends upon certain underlying neural events, but not necessarily upon a particular location of the excitation in either the retina or in area 17 of the visual cortex. Throughout this paper "phenomenal location" (in contradistinction to "anatomical location") will be used in this sense.

The scene consisted of two vertical luminous lines on an otherwise dark homogeneous field. About 22 in. in front of the artificial pupils was the arrangement which produced the vertical lines. This consisted of a ground-glass screen upon which were mounted two pieces of black cardboard, each containing a vertical slit. (The slits were 3/16 in. in thickness and 3 in. in length.) The cardboard pieces, each containing a vertical slit, were mounted on a track over which they could be shifted horizontally behind the ground-glass screen so as to increase the distance between them when desired. Only one line was visible through each aperture and it was only visible when the eye was turned to fixate it through the aperture. Thus, when *O* viewed through the left aperture, only the left line was visible and when *O* viewed through the right aperture, only the right line was visible.

Extreme care was taken to insure that each line stimulated only the foveal portion of the retina. This was accomplished with each *O* in advance of the experiment by illuminating the aperture and the lines at the same time. *E* then moved the lines until their images were as close to the inner or nasal portion of each aperture as possible, while *O* fixated the lines. This insured that the lines would disappear from *O*'s view with the onset of the slightest lateral eye-movement in a nasal direction. With the eye so close to the aperture, any turning brings the pupillary opening out of alignment with the entire cone of light rays coming through the aperture from the line. It was not possible with the arrangement used here simultaneously to prevent *O* from seeing a given line peripherally through an aperture by turning his eye beyond the line. Under instructions to fixate, it is unlikely, however that this would occur. If it were to occur a theory based on neural interaction would have to predict movement in the opposite direction to that which was experienced based on an analysis of the proximal stimulus-situation. (See the discussion of this problem p. 203.) In other words, *O* could see each line by fixating it and nothing else was visible but the fixated line. As soon as the vertical lines were in place, the surrounding area was so darkened that the apertures were themselves no longer visible to *O*, but of course he could still see *through* them. By means of a dual projector and an episcotister, the two vertical lines were made to appear alternately, one through each aperture. *O* was then instructed to move his eyes from left to right, etc., so that the illuminated vertical lines would be in foveal view at each position of the eye. Fig. 1 illustrates the experimental arrangement. All *O*s, after a few trials, were able to synchronize their eye-movements to the alternating vertical lines.[10] *O* was then asked to report his spontaneous impression of the scene. The *O*s were students at the New School for Social Research; some of them probably had seen stroboscopic movement demonstrated in a class, but all of them were naïve about the hypothesis. Care was taken not to mention that the experiment had to do with stroboscopic movement.

---

[10] The main purpose of using the alternating projection of the vertical slits was to force *O* to achieve a particular speed of alternation of eye-movement. No doubt a stroboscopic effect would have been achieved even with both slits continuously illuminated because only one is visible at a time, but then there would be no control over rate of eye-movement and, hence, over rate of alternation. It is, however, possible that our arrangement did also contribute to a flashing on-and-off effect. Without it, *O* might have in some way sensed that it was his own eye-movement which 'turned off' a line and, if so, the lines may have been seen as permanently on but not always visible, which probably is not an ideal condition for producing a movement-effect.

The precise conditions of rate of alternation and distance between lines varied somewhat for each *O*, depending upon the rate of eye-movement he was able to attain. For one *O* believed to be representative, the rate of alternation was approximately 0.3 sec. with the lines set 7 in. apart. Each line remained exposed for approximately 0.13 sec.

In summary, the major point of the above procedure was to insure (1) that the proximal stimuli would fall in a single retinal region, and (2) that the corresponding distal stimuli would be experienced at different phenomenal locations, once to the left and once to the right. Such phenomenal localization is based on the change in eye-position.

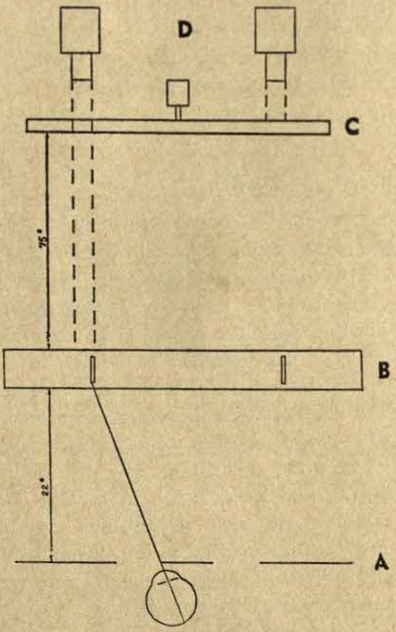


Fig. 1. Schematic Diagram of Arrangement Used in Experiment I, Part 1
A, artificial pupils; B, ground-glass screen, with vertical slits in cardboard; C, episcotister in position of transmitting light from left projector; D, dual 300-w. projector.

Results. Of 10 *O*s who succeeded in properly coördinating their eye-movements to the alternating stimulus, 6 reported seeing a single object in motion almost immediately following the onset of stimulation. The remaining 4 *O*s reported seeing 2 objects, both of which were flashing on and off in their respective positions. These same 4 *O*s were, however, unable to perceive apparent movement under normal viewing conditions.[11] These cases do not, therefore, represent negative instances of the hypothesis in question. The results thus far indicate that stroboscopic motion can be experienced in the absence of disparate retinal-cortical stimulation.

Part 2. In Part 2 of this experiment, the attempt was made to create a situation in which different regions of retinal stimulation were phenomenally localized in the *same* region in space. In this situation, one eye-position allowed for foveal stimulation of the object whereas the other eye-position allowed for peripheral stimulation by the same object while fixating a small stationary spot. In both cases, however, the stimulus-object was experienced at the same locus in space. This was accomplished by requiring *O* to move his eye back and forth so as to view a single stationary flashing line alternately foveally and peripherally.

Procedure. As soon as Part 1 was completed, one of the vertical lines was removed and a hole ½-in. in diameter was substituted for it. This hole was continuously illuminated and served as a fixation-point. The two projectors behind the episcotister were now both trained on the same vertical slit from behind the ground-glass screen. This arrangement was thus one which illuminated the vertical opening at a rate exactly twice that of Part 1. (The reason for the increased rate was that every other exposure was intended to correspond to the exposure of a second line. Thus, the rate between exposures of the line equalled that between exposures of the two lines in Part 1.) The scene, therefore, consisted of a line flashing on and off and a continuously illuminated fixation-point. *O* was then asked to view the scene monocularly through a circular hole cut in black cloth (placed about 3 in. above the artificial pupils used in Part 1). The opening was fitted with two polarizing filters at such an angle with respect to each other as to maintain a black homogeneous surround for the flashing line and the fixation-point. *O* was instructed so to adjust his gaze between the line and the point that the line would stimulate alternately the foveal and peripheral regions of the eye with each flash. This was accomplished by the following succession of eye-movements: *O* fixates point on right (line flashes in peripheral vision on left); *O* fixates line on left (line stimulates fovea); *O* fixates point on right (line flashes in peripheral vision on left) etc. Thus, as far as the retina or cortex is concerned, there is alternate stimulation of disparate points exactly as occurs in a more typical demonstration of apparent movement. The major difference, however, is that here the flashing source will be localized at one and the same place in space. After several trials, all *O*s were asked to report their spontaneous impressions of the scene.

---

[11] They were given the opportunity to view the alternating stimuli monocularly and without an artificial pupil. The rate of alternation was kept identical with the highest rate of eye-movement attained in the experiment proper. The reason these *O*s did not experience movement may be that their threshold for stroboscopic motion required a rate of alternation well above that at which they were capable of maintaining lateral eye-movements.

Results. The six *O*s used in this part were the same *O*s who experienced movement in Part 1. Of the six, none reported the perception of movement under these conditions.

These results indicate that in the absence of divergent spatial locations of the distal stimuli, stroboscopic motion will *not* occur, even though disparate retinal points are being stimulated. Apparently the fact that the line is experienced in one and the same place in space eliminates the possibility of seeing movement.

EXPERIMENT II

Part 1. As in Experiment I, Part 1, the over-all intent of Experiment II was to create the conditions under which an object would stimulate a given (constant) portion of the retina, but be phenomenally localized at two different places in space. The two experiments differ, however, in the way in which the disparate object-localization was achieved. In Experiment I, change in eye-position may be assumed to have been the major determinant of change in phenomenal location. In Experiment II, it was decided to utilize as the determinant the change in relative position of one object with respect to another.

Procedure. The conditions employed are quite similar to those of Duncker's demonstration of induced motion.[12] In Duncker's demonstration, motion was attributed to an object that had no moving counterpart in the retinal pattern. He attributed this effect to what he termed object-relative displacement. In the case where one object may be said to surround another or to act as a frame of reference, when the latter (the frame of reference) is set in motion, the enclosed, surrounded object is characteristically perceived as moving in a direction opposite to its frame.

The present conditions differ from those of Duncker with respect to the fact that the object expected to be seen as moving is not continuously present but instead flashes on and off in the same physical locus. A luminous line was moved back and forth (frame of reference), while a shorter luminous line flashed on and off in a fixed position in space. The short line thus appeared now to the left of the larger moving line, now to the right.

O was at a distance of about 4 ft. from a display which moved in a plane perpendicular to his line of sight. He viewed binocularly, in a completely darkened room, from a fixed head-position. The display consisted of the following: In the center of a black cardboard was a vertical luminous line of ½-in. width and 8-in. length. In symmetrical positions, to the left and right of this line, were two slits, ⅛ by 2 in. long. The slits were exactly 2⅝ in. apart at their extreme edges.

The entire display was mounted on a wooden frame which moved in a track laterally, for a total distance of 2⅝ in. Upon reaching the point of maximal displacement in any one direction, the display immediately changed its course and proceeded in the opposite direction. The speed of alternation could be varied by means of variable rheostat. The movement was accomplished by means of a reduction motor which drove a circular metal disk in a horizontal plane. A shaft mounted at a point on the perimeter of the disk was attached to the wooden frame. As the metal disk revolved, a force was applied to the frame via the shaft and the ensuing direction of movement was determined by the horizontal track. The total displacement of the display (2⅝ in.) was determined by the diameter of the metal disk (and angle of disk to track).

---

[12] Karl Duncker, Über induzierte Bewegung, *Psychol. Forsch.*, 12, 1929, 180-259.

At a fixed point on a wall directly behind the display was a single luminous line of dimensions identical with one of the slits in the display. This line was so situated as to coincide with each slit at the two points of maximal displacement of the display. Thus, as the display reached its left terminal point, the right slit became coincident with the luminous line; as the display reached its right terminal point, the left slit coincided with the same luminous line. This situation provided, therefore, for the alternate illumination of the left and right slits, although the source of illumination was at a fixed point in space.

O was instructed to fixate a luminous spot fastened to a clear glass plate at a distance of about 2 in. in front of the display and placed directly in *O's* line of sight. Under these conditions, the stimulation (coming alternately from each slit and provided by the fixed luminous line behind the display) consistently fell on the same portion of the retina. Thus as in Experiment I, Part 1, the classical conditions for phi (viz. the alternate stimulation of disparate retinal points) were not present. The question was whether the possible shift in phenomenal location of the flashing line induced by the actual change in position of the larger line would yield the impression of movement.

When the apparatus was set in motion, *O* was required to report what appeared to be taking place in his field of view. Different *O*s participated in this variation.

Results. All of the *O*s (10) who took part reported viewing a single short luminous line describing a semicircular path about a taller luminous line in the third dimension. The center line was also seen to move back and forth. About as many *O*s reported the motion of the short line taking place behind the tall center line as did those reporting motion in front of the line. The results clearly support the hypothesis that change in phenomenal location without change of retinal location will yield the perception of stroboscopic motion.

This procedure on the surface is similar to one used by Duncker whereby stroboscopic motion can be used to illustrate induced movement. The relative position of an outer rectangle changes in two slides, alternately projected, while that of a central dot remains unchanged. The rectangle is seen to move but so is the dot, despite the fact that the dot's position is objectively unchanged. We would have cited this demonstration to support the proposition of the dependence of stroboscopic movement on change of phenomenal location rather than retinal location (instead of performing the above experiment) were it not for one flaw. The retinal position of the rectangle *does* change so that one might argue that the stroboscopic movement induced in the dot is transferred in some way from the stroboscopic movement of the rectangle and *that* movement *is* based on change of retinal location. In the present experiment, however, the central line is in *continuous* (real) movement and the short line merely flashes on and off in the same objective place. There is thus no stroboscopic motion to be transferred to the short line; nevertheless, a stroboscopic effect is achieved.

Part 2. The results of the previous experiment indicate that a discontinuous change in the relative position of the stimulus-object with respect to another object was sufficient to produce stroboscopic motion. This was the case despite the absence of disparate retinal stimulation.

If we consider this experiment from the perspective of Duncker's conceptual system, it is probable that the constantly illuminated center line served as a frame of reference, thereby 'inducing' stroboscopic motion in the alternately illuminated line. If Duncker's formulation can be applied to the present instance, then if the object, with respect to which a change in position takes place, were made relatively small as compared with the alternating stimuli, we should expect no induced movement because it would no longer have the character of frame of reference. The question remains, therefore, as to whether the experience of motion can be eliminated by changing the characteristic of the flashing stimulus to give *it* the property of frame of reference.

Procedure. The display of Experiment II, Part 1, was modified by making the slits 6 in. long. The luminous line, with respect to which the alternation occurred, was reduced to a tiny luminous spot (¼ in. square). The same apparatus and method of illumination was used as in the first part of Experiment II. Thus, the left and right slits were alternately illuminated from a fixed source behind the display.

Five naïve *O*s and five *O*s who were sophisticated about stroboscopic movement, but not the hypothesis under investigation, were used. They were given the instructions of Experiment II, Part 1.

Results. None of the 10 *O*s reported any motion of the vertical line. All indicated that the same line was going on and off at the same point in space. Of course, the small spot actually moving back and forth was seen to do so.

Part 3. In Part 3 an attempt was made to utilize Duncker's principles to interfere with stroboscopic movement under conditions where it might be expected to occur. It was required that the object, which now appeared in a constant position relative to another object, stimulate disparate portions of the retina. Thus, although the relative position of the stimulus-object remained constant, the corresponding retinal stimulation alternated between two points. Here, as in Experiment I, Part 2, we have represented, therefore, the classical conditions for stroboscopic motion.

Whereas in that experiment a change in retinal location was accompanied by a constant phenomenal location, in the present experiment it might be thought that the change in retinal location is accompanied by a change in phenomenal location. Insofar, however, as relative position is also a determinant of phenomenal location, and insofar as the latter does *not* change, this experiment is, in essence, putting in conflict two systems of cues to phenomenal location.

Procedure. In this part the apparatus remained the same as in Part 1 of Experiment II. The display, however, was modified in two ways. First, the left slit was covered over to block out any illumination from behind. Secondly, a second luminous strip of the same dimensions as the one used in Part 1 was mounted on the wall behind the display. This was positioned in such a manner as to correspond with the right slit at the extreme point of its left movement; and, as in Part 1, the right slit was illuminated at the extreme point of its right movement. Thus, the same slit was illuminated alternately at the two extreme points of its lateral displacement. The long center line remained continuously in view as it moved, as in Part 1. The outcome of this arrangement was that an 8-in. line moved back and forth and a 2-in. line flashed on 2⅝ in. to the right of the longer line at each of the terminal positions of the longer line. *O* was instructed to fixate a luminous point in front of the display as in Part 1 and to describe the visual scene.

Results. The *O*s of Part 1 were also used for Part 3. All *O*s found great difficulty in describing their experience. A typical report was as follows: "The small line seems to be moving and at the same time standing still." It will be recalled that in this condition, two systems of cues which normally act jointly to determine the spatial location of objects, are here opposed. That is, normally a change in retinal location with eye- and head-movement eliminated, will indicate a change in phenomenal location. On the other hand, the maintenance of a constant position relative to some external frame of reference serves to indicate a constant phenomenal location within that system. In light of this analysis the ambiguity expressed in the *O*'s reports is readily understandable.

On the whole, the results of this experiment seem to imply that to obtain the unequivocal experience of stroboscopic motion, change in phenomenal location must itself be unequivocally given. That is, when stimulus-conditions are such that phenomenal location is ambiguous, so too is the resultant stroboscopic movement experience.

DISCUSSION

One criticism that at an earlier time might have been made against Experiment I, Part 1 (undoubtedly our most important experiment) is that the cause of the experienced motion can be attributed to the eye-movements. By now, of course, it is clearly established that eye-movement is not a necessary condition for stroboscopic movement[13] and no one has ever demonstrated that it is a sufficient condition. We believe, therefore, that we are free to use change of eye-position as a technique for bringing about change of phenomenal location. In Experiment I, Part 2, the same eye-movements do not yield any experience of motion. Also it may be noted that Experiment II, Part 1 does not entail change of eye-position as an independent variable.

There are several facts about stroboscopic movement that support our over-all conclusions.

(1) A theory of stroboscopic motion based on neural interaction could never, in itself, be a complete theory for the following reason: Logically it is possible that the interaction could explain why an experience of movement occurs but it does not adequately deal with the problem of the experienced directions in space of the starting and ending positions of that movement or, in short, its path. If location were purely a matter of the locus of the proximal stimulus (as is no doubt the case in species whose eyes do not move), then the 'path' of the neural interaction would correspond with the path of the seen movement in space. Where, however, eye-position plays a central role in the location of objects with respect to the perceiver, certain logical difficulties arise. Suppose *O* does move his eyes somewhat while viewing the flashing points. Suppose, for example, he fixates one stimulus and fixates only slightly to the side of the other. If he sees movement, as we have every reason to assume he will, it will not merely be a slight movement, as might be predicted from the slight retinal shift involved in our example. Rather he will see the movement between the two points as veridically localized in phenomenal space, each on the basis of eye-position together with retinal position. Hence, the interaction between the anatomical loci of the stimuli *per se* does not tell us anything about the path of motion. That path corresponds with the

---

[13] Wertheimer, *op. cit.*, 161-278; J. P. Guilford and Harry Helson, Eye-movements and the phi-phenomenon, this JOURNAL, 41, 1929, 595-606.

phenomenal location of the terminal positions and not necessarily with the retinal (or cortical) terminal positions. From this analysis, our experiments, in which no shift in the retinal stimulus occurs at all, represent the extreme case and show in addition that no interaction is necessary for the movement experience.

A related fact is that stroboscopic motion occurs in the third dimension of space under conditions where the source is phenomenally located at different distances from *O*. This illustrates the point made above; namely, the path of the movement does not correspond in any direct way with the path one would predict based on retinal stimulation—*i.e.* the 'path' of the alleged interaction between the anatomical loci of excitation.

(2) After these experiments were completed, a paper by Verhoeff which reports on certain clinical observations bearing on the present problem came to our attention.[14] Verhoeff demonstrated that stroboscopic movement can be elicited by alternate stimulation of the two foveas of squinters with anomalous projection. In such patients the image on one fovea is generally localized separately from that on the other. Thus, here, too, we see that the movement experience occurs so long as the light-sources are localized separately in space, despite the fact that only one retinal (and cortical) region is stimulated.

(3) The effect has been obtained with interocular stimulation, where the two excitations emanate from different eyes and end in different hemispheres of the brain.[15] This condition would seem to preclude peripheral interaction and to make unlikely central neural interaction. The effect has also been obtained across acquired scotomata of cerebral or retinal origin.[16]

(4) It has long been known, as one of Korte's Laws,[17] that as the distance between sources is increased, the time-interval between exposures must also be increased if an effect of movement is to continue to occur. The question arises, however, whether distance here is to be conceived of in retinal or in phenomenal terms. Corbin showed that Korte's effect occurs when the distance is so increased along a plane inclined to *O* as to hold constant the retinal distance.[18] If stroboscopic movement does depend on stimulus-conditions which yield change of phenomenal location as we have

---

[14] F. H. Verhoeff, Phi phenomenon and anomalous projection, *Arch. Opthal.*, 24, 1940, 247-251.

[15] K. R. Smith, Visual apparent movement in the absence of neural interaction, this JOURNAL, 51, 1948, 73-78.

[16] H. L. Teuber and M. B. Bender, Perception of apparent movement across acquired scotomata in the visual field, *Amer. Psychologist*, 5, 1950, 271 (abstract).

[17] Adolf Korte, Kinematoskopische Untersuchugen, *Z. Psychol.*, 72, 1915, 193-206.

[18] H. H. Corbin, The perception of grouping and apparent movement in visual depth, *Arch. Psychol.*, 38, 1942, (No. 273), 1-50.

argued, then, insofar as distance enters into the process, we would expect it to be phenomenal distance, which apparently is the case.

(5) There is considerable evidence that set, experience with the effect, and extent of training influence the perception of stroboscopic motion.[19] Such facts do not seem to follow from neural interaction theories which must necessarily regard the effect as autochthonously determined.

There are both negative and positive implications of our finding. On the negative side, it seems to follow that all theories which seek to explain the experience of stroboscopic movement in terms of some neural interaction between the loci of excitation are in error. A theory of interaction is only appropriate if it is correct that a necessary condition is stimulation of anatomically separate regions, and we have shown that this is not a necessary condition. In Gestalt thinking such a theory exemplified the postulate of isomorphism—viz. if motion is seen between two points in space, then some neural correlate of motion must exist between the cortical excitations representing those points in space. Without necessarily challenging isomorphism as such, our findings do challenge the way isomorphism has been applied in perception; namely, the seeking of correlates of various aspects of phenomenal space in terms of interactions within the cortical space. In other words, facts of space-perception cannot be accounted for purely in terms of a theory tailored to the *spatial* dimensions of the visual cortex. If, for example, only a single cortical region is excited, it can nevertheless represent various points in phenomenal space, depending on eye-position. In principle, all of space could be represented by a single retinal (and cortical) point, thus showing that perceived spatiality does not necessarily correspond with cortical spatiality in any simple or direct fashion.[20] This is not to deny, of course, that perceived spatiality corresponds with some cortical events which represent space. In the long run, an appropriate neural explanation for the experience of stroboscopic movement will be found—which may or may not be compatible with the doctrine of isomorphism—but it will undoubtedly be quite different from present theories. We suspect, in line with the discussion below, that such an explanation will be similar to that which will be found to explain all perceived movement.

---

[19] C. I. Hovland, Apparent movement, *Psychol. Bull.*, 32, 1935, 755-778; W. S. Neff, A critical investigation of the visual apprehension of movement, this JOURNAL, 48, 1936, 1-42; E. E. Jones and J. S. Bruner, Expectancy in apparent visual movement, *Brit. J. Psychol.*, 45, 1954, 157-165; H. E. Toch and W. H. Ittelson, The role of past experience in apparent movement: re-evaluation, *ibid.*, 47, 1956, 195-207; William Epstein and Irvin Rock, Perceptual set as an artifact of recency, this JOURNAL, 73, 1960, 214-228.

[20] We do believe that the spatial dimensions of the cortex correspond with the experienced (two dimensional) topological relations of the momentary visual field. Perhaps all original experiences of extensity, shape, relative position, and the like depend upon the spatiality of the visual cortex. Nevertheless the stable visual world involving the location of things with respect to the observer is also very much a function of eye-position, *i.e.* of where the momentary field is 'aimed.' Thus objects stimulating, say, the fovea, although always located in the center of the momentary field, are located in space with respect to the self in whatever direction the eye is turned. (Cf. J. J. Gibson, *The perception of the visual world,* 1950, 26-43.) The integration of retinal position with eye-position in determining phenomenally perceived direction may be a matter of development and learning, at least in humans.

Somewhat similar reasoning applies to the peripheral theories of interaction. Only if it were the case that phenomenal space corresponds in a direct fashion with retinal 'space,' could such theories logically be considered tenable. Whatever, therefore, may be the evidence for neural interaction in the retina, our evidence shows that it cannot be considered a necessary condition for stroboscopic movement.

On the positive side, our finding that stroboscopic movement will only be experienced when change of phenomenal location is involved dovetails nicely with certain facts about the perception of movement in general. In the case of fixating a moving object, it is also clear that it will be perceived as moving, if its displacement is above threshold, despite the fact that here, too, there is no retinal (or cortical) displacement. (Cf. the analogous Experiment I, Part 1.) Of course, the converse, sweeping the eyes past a stationary object, does not yield an impression of movement despite the displacement of the image. (Cf. the analogous Experiment I, Part 2.)[21]

In the case of induced motion, the stationary object is seen as moving and, if below threshold, the moving object as stationary. Here, again, it would seem that change of phenomenal location occurs. The induced object is seen as changing its location with respect to the inducing one which acts as a frame of reference. (Cf. the analogous Experiment II, Part 1.)

Our findings also fit in with what is now known about phenomenal velocity, which is the quantitative aspect of phenomenal movement. Brown showed that perceived speed is a function not of the absolute rate at which an object's image traverses the retina but rather of the rate at which it displaces relative to its frame of reference.[22] In other words, we may say that phenomenal velocity is a function of rate of change of phenomenal location, not of retinal location. The fact of speed-constancy also supports this conclusion.[23]

---

[21] In the case of real movement, Gestalt psychologists, as for example Duncker or Koffka, have been well aware of these facts, but curiously they failed to see that they might equally be true of stroboscopic movement, and that, if so, it would contradict Wertheimer's type of theory.

[22] J. F. Brown, The visual perception of velocity, *Psychol. Forsch.*, 14, 1931, 199-232.

It therefore seems correct to generalize that the perception of movement depends on those stimulus-conditions which can yield an experience of change of an object's phenomenal location, providing, of course, that the change is above some threshold-value.[24] There is a certain logic to this, since, in reality, if an object changes its location, it does so by moving. (The focus of interest, therefore, shifts to the problem of uncovering the precise determinants of phenomenal location and to the problem of threshold.) Ultimately, an explanation in terms of brain-process will be one tailored to deal with this central fact.

From this point of view, there is only one remaining problem about stroboscopic movement that in any way makes it a special case. That is the fact of discontinuous rather than continuous stimulation. Some authors have tried to argue that this is not a crucial distinction, since the retina consists of cell which, even if adjacent, also are discontinuous. Hence, even real movement involves discontinuous stimulation, only less so. But as shown by the results reported in this paper, the discontinuity should not be thought of in retinal terms at all but rather in spatial terms. Even a flashing on and off in the *same* retinal location will, under the proper conditions, yield an impression of movement across space. In these terms, the problem *does* remain of why movement is perceived across a rather substantial stretch of space when there is no stimulation corresponding to continuous displacement.

Once it is realized, however, that stroboscopic movement can be subsumed under the general principle of change of phenomenal location, this phenomenon is perhaps no longer so puzzling. In line with the 'logic' that perception often manifests, if the identical thing is now 'here' and now 'there,' then it can only have changed position by moving. (The notion of identity has been stressed by many investigators of stroboscopic movement and the similarity of the two exposed objects is now known to be important.) Furthermore the stimulus-conditions are quite similar to those which obtain during real movement. The terminal positions of a really moving object perhaps form the more important component of the stimulus-conditions, particularly if the object moves rapidly, since the intervening positions are often blurred out or not noticeable as such. From this point of view, one might predict the phenomenon of stroboscopic movement even if it were not known, based on similarity—a kind of stimulus-generalization.[25] It, therefore, becomes at least plausible that the effect is a function of past experience with real movement.

---

[23] Hans Wallach, On constancy of visual speed, *Psychol. Rev.*, 46, 1939, 541-552.

[24] Below threshold, a change in position will be noted but it will not be accompanied by the quality of movement, as for example the minute hand of a watch.

SUMMARY

The question was raised whether stroboscopic movement depends upon successive stimulation of separate retinal (and therefore cortical) points or of points located separately in phenomenal space. Ordinarily the latter requires the former, but it is possible to create conditions where phenomenal separateness is experienced even when only one region of the retina is stimulated. With the use of two different techniques, it was shown that under such conditions stroboscopic movement is experienced. Conversely, it is possible to stimulate separate retinal points in such a manner that the source will not be experienced in two localities but, rather, in only one. It was shown that under such conditions movement will not be experienced. These findings were taken to imply that neural interaction between two loci of excitation (on any level) cannot be a general explanation of stroboscopic movement. It was pointed out that the findings are consistent with certain facts concerning movement-perception in general; namely, that motion is experienced whenever above-threshold change in the phenomenal location of the source occurs.

---

[25] Cf. R. S. Woodworth and Harold Schlosberg, *Experimental Psychology*, 1954, 515.

THE DIRECT MEASUREMENT OF UTILITY AND SUBJECTIVE PROBABILITY

By EUGENE GALANTER, University of Pennsylvania

The recent experimental work on choice stresses either one of two general concepts of behavior. One is learning theory in which the units of behavior are responses. These constitute the elements among which an organism chooses, and the termination of a response results in the occurrence of certain events, generally called reinforcers. The choice among responses is presumably determined by sequences of previous responses and the reinforcements contingent upon them. Often the contingencies between response and reinforcement are treated as probabilistic, and much work has gone into an analysis of schedules of reinforcement.[1]

Recently, psychologists have been presented with an alternative conceptualization of behavior that stems from the work of economists, statisticians, and mathematicians.[2] This conceptualization has many parallels to the previous description, but in the place of responses, it substitutes courses-of-action as the units of behavior. A choice is made among a set of alternative courses-of-action. Once a course-of-action has been selected, it is presumed that certain events occur that are to be called 'outcomes.' More commonly than is the case with reinforcement, an outcome bears a probabilistic rather than an uniquely determined relation to its antecedent course-of-action. If the objective probabilities of the outcomes are known, the choice between alternatives is termed 'risky.' If the objective probabilities of the outcomes are unknown, the choice is termed 'uncertain.' Whether the choice is risky or uncertain, the degree of belief about the outcome is a subjective event that must be measured psychologically. This measure is often called a 'scale of subjective probability.' The question of whether these subjective probabilities exist can be answered only in a logico-empirical context. That is to say, if, given the data from a variety of experiments on choice and a theory summarizes the data, then, if the theory contains as a variable a scale whose natural interpretation is that of a sub-

---

* Received for publication January 27, 1959. Supported in part by the United States Air Force under Contract No. AF 41 (657)-118. George Letchworth helped in various ways in the conduct of these experiments.

[1] Charles Fester and B. F. Skinner, *Schedules of Reinforcement*, 1957, 326.

[2] Reviewed by Ward Edwards, The theory of decision making, *Psychol. Bull.*, 51, 1954, 380-413.

jective probability, for that theory the existence of the scale is guaranteed.

The concept of subjective probability serves the same purpose for the choice-theorist as does the idea of partial reinforcement for the psychologist using learning theory. Further than this, however, a theory of choice is concerned with outcomes that have a desirable and undesirable property as a function of the evaluative processes of the organism. This property, analogous to 'incentive value,' is a function on outcomes that assigns to each outcome a numerical quantity termed its 'utility.' By the use of utility as a defined measure of outcomes, economists study what the behavior in a choice-situation *should* be. That is to say, the economist selects some function of utility and some mathematical operation like maximization, minimization, or equilibration. The combination of these terms defines 'rational' behavior. He can then test the theory.

There have been many suggestions about these normative modes of behavior, and a few of the so-called rational criteria of choice have been investigated experimentally.[3] It is characteristic of almost all the criteria, however, that the variables entering into the choice among alternatives are more or less exhausted by two numbers, specifically the likelihood that a particular outcome will occur when any given alternative is chosen, and the utility of the outcome given that it does occur. The essential difference between the theory of choice and the theory of learning is that the choice-experiments consistently contain two independent variables instead of one. These are (1) the utility of the outcomes and (2) the likelihood of the outcomes, and little attention need be given to the sequential aspects of the behavior. The learning theorist, on the other hand, usually controls the utility of the outcome by using a constant reinforcement. He can then study either the effects of the various probabilities of reinforcement and how they determine the *course* of choice, or the preferences of an organism for various outcomes. The choice-theorists' way of looking at behavior is in one respect like the psychophysical experiment. In that kind of an experiment the scientist varies the stimuli and observes the changes in the probability of choice; it is unlike the learning experiment where the stimuli are kept fixed and only the sequential features of the choice are recorded.

The psychophysical, or static, study of choices should eventually interlock with the dynamic, or learning, study of choices. A theory of choice in which both the static and dynamic aspects are treated as two functions of the same mechanism is exemplified in Luce's recent book.[4] Work in modern psychophysics that uses differential payoffs as outcomes for the psychophysical judgments is another example of the rapprochement between the work in learning and in psychophysics. The psychophysical function becomes the distribution function of the asymptote of a set of parallel curves from experiments in discrimination learning.

---

[3] Donald Davidson, Patrick Suppes, and Sidney Siegel, *Decision Making,* 1957, 1, 9 ff.

The program set forth by a theory of choice is, of course, very enticing. This program contains elements that are surely important in the determination of human choice, and this scheme represents a single choice-situation in terms more general than the simple T-maze. Unless, however, the theory is able to specify how we shall measure independently the utility of the outcome and its likelihood, there is no way of predicting the choice. Those men who have elaborated the theory are well aware of their problem, and they have devoted much effort to finding ways to decompose individual choices into their component utilities and likelihoods.

In general, assumptions are made about how a 'sensible' person would behave, and then, by assuming that an experimental *S* did behave that way, the numerical values of one or the other or both parameters are inferred.[5] The particular methods of decomposing choices into these two components involve both mathematical and psychological assumptions. The decomposition is normally effected by introducing axioms about what people will do. Thus, to measure the utility and the likelihood of the courses-of-action, the choice theorist presumes, or perhaps creates, a theory of psychology. It would be preferable to measure the utility of certain outcomes independently of the choice, and measure the likelihood of the outcome also independently of the choice. With these two scales, it would be possible to discover the psychological rules that effect their combination.

The present study was designed, therefore, to discover whether direct measurements of utility and likelihood are possible. The study divides itself into three parts: (1) The measurement of utility; (2) the measurement of likelihood by three methods, and (3) the prediction of choices made by people when we have knowledge of the likelihood and the utility of the alternatives.

THE DIRECT MEASUREMENT OF UTILITY

With many sensory continua, a number of techniques have been devised for the construction of scales of subjective magnitudes. All of these techniques depend in some measure upon the inability of the *S* to estimate the physical magnitude of the stimulus precisely. The stimuli that are to be scaled should be confusable so as to minimize the so-called stimulus-error.

---

[4] R. D. Luce, *Individual Choice Behavior: A Theoretical Analysis*, 1959, cf. 38, 91.

[5] Davidson, Suppes, and Siegel, *op. cit.*, 35.

If one selects a stimulus-continuum that has cardinal numbers as names for the elements of the continuum, as money does, the possibility of scaling that continuum in isolation seems remote. The problem is that if you ask people to tell you what half of $20.00 is, they will always say "ten dollars." It is obviously difficult to get people to detach themselves from the numerical nomenclature associated with money and make them express the value or utility of the money to them; but the advantage of scaling the utility of money is the clear-cut unidimensionality of value that this commodity possesses. Since this continuum is one over which we can effect some control, it it worth trying far-fetched techniques if they offer any hope.

The problem is, therefore, to get people to scale not amounts of money but, rather, the valuational effect that the receipt of money produces. With this scale numerical symbols are not normally associated. To this end people were asked to estimate the "happiness that money brings."

Twenty *S*s were given the following instructions:

Suppose I were to give you, as an outright gift, $10.00. This $10.00 does not come from me, but comes from a foundation whose resources are limitless. You will be taking nothing away from me when I give you this $10.00. Presumably, this will make you happy to a certain degree. Now, I want you to think about how much money you would want in order to feel twice as happy as the $10.00 would make you feel."

This set of instructions consistently evoked responses about the impossibility of performing the task, but *E* was insistent and prodded *S* into making a thoughtful estimate. All the *S*s finally made an estimate. The same procedure was repeated with two additional groups of 20 *S*s each, except that for the second group, the base amount of money was set at $100.00, and for the third group the base amount of money was set at $1000.00.

Our question now is: What additional increment does the typical *S* require to feel that the happiness associated with the gift would be doubled? To determine this increment, we computed as a measure of location both the means and the medians of the 20 judgments in each of the three situations. These means and medians are listed in Table I. Table I also contains the range of the judgments. The distributions were skewed because of the kind of natural lower limit that twice the amount represents. In only 3 of the 60 cases did any *S* want less than twice the amount of money in order to be twice as happy.

The first observation that we can make is that the relative magnitude of the increase seems roughly comparable for each of the three widely divergent levels of money with which we dealt. The second observation is that there is a good deal of scatter even in these measures of location. Any methods that we can suggest for striking an average among these diverse values will be, at best, a makeshift one. How, then, to proceed?

The scale of utility that we accept can be relatively arbitrary because its ultimate value will depend upon our ability to predict certain choices among amounts of money. We want, however, to establish a scale that has some intuitive appeal. Another criterion is the relative simplicity of the function that we use to describe utility, and the ubiquitousness of the function among other quantitative perceptual continua. If we make the reasonable conjecture that constant ratios of money yield constant ratios of utility, then we assert that the function relating utility to money is a power function, of the form $U = kM^n$. This form is identical with ratio-scales for 14 other analogous perceptual continua.[6] The critical parameter in this function is the slope parameter, n. As a best guess, we have selected the ratio of five to two as the ratio of the rates at which money and utility grow. By this we mean that, to obtain twice the utility, you have to get five times the money. This sets the value of the exponent, n, at 0.43. If we then define the units of utility (the utile) by making $10 equal 10 utiles, the coefficient, k, is set equal to 3.71. Fig. 1 presents the relation between money and utility. Observe that the utility-function has the additional, intuitively appealing property of a decreasing marginal utility. Although one might be able to invent examples in which marginal utility increased, it is certainly reasonable that as a first approximation the function (in arithmetic coördinates) should be concave downward.

TABLE I

AMOUNT GIVEN AND AMOUNT REQUIRED TO DOUBLE IT

Given	Mean	Median	Range
$10.00	53.00	45.00	20.00– 100.00
$100.00	537.50	350.00	150.00– 800.00
$1000.00	10220.00	5000.00	3200.00–16000.00

This result requires two qualifications. In the first place, we have asked people about *hypothetical* increments of money, and we do not have any information about what would happen if they *actually* got the money. This is often stated as an argument against techniques of this kind; but the fact of the matter is that prior to a decision one always considers the alternatives as hypotheticals and, presumably, bases his action on these considerations. To argue that we use an hypothetical situation to scale the value of money and, therefore, that the scale does not represent what people do, is to prejudge the usefulness of the scale as a characterization of the *hypotheses* that antedate the decisions. It remains to be seen whether this scale about hypotheticals can predict actual choices.

---

[6] See S. S. Stevens, On the psychophysical law, *Psychol. Rev.*, 64, 1957, 153-181; S. S. Stevens and E. H. Galanter, Ratio scales and category scales for a dozen perceptual continua, *J. exp. Psychol.*, 54, 1957, 377-411.

The second point deals with the question of assessing the utility of decrements in money. Over a period of time we have tried to get people to tell us how much of a loss would make them twice as unhappy as, say, losing $100.00. We have had no success. People find the question much more difficult than the incremental question. They often give a number of answers and tell us to choose the one we mean. So far we have been unable to get a relatively stable answer to the question. It appears, from the scanty data at hand, that the utility-function for negative amounts of gain is most likely a power larger than two of the money. If this can be shown it would provide a simple explanation for certain common phenomena like 'Indian giving,' or Machiavelli's dictum to the Prince to make losses to the people in a single installment.

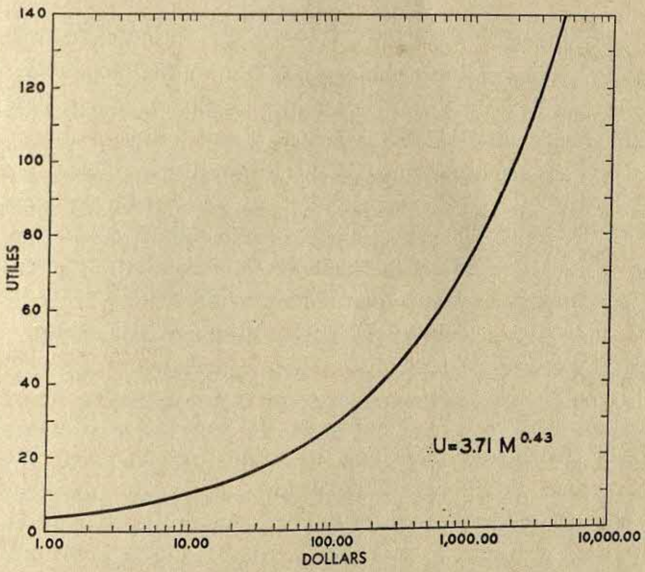


FIG. 1. THE UTILITY-FUNCTION FOR MONEY
The ordinate represents the subjective utility of money measured in arbitrary units called 'utiles.' The multiplicative constant of the function was so selected that $10.00 equals 10 utiles. The abscissa is a logarithmic scale of money. In an arithmetical plot the function would be concave downward.

The analysis of these data assumes that all of our *S*s were at zero utility when we questioned them. Is this assumption tenable? Probable not in detail, but as a rough approximation it seems safe to assume that *S*'s capital did not seriously affect his judgments.

We must now ask about comparable results if our question is rephrased in terms of halving instead of doubling. Fourteen new *S*s were asked how much money would make them half as happy as $100.00. Twelve of the *S*s said $50.00. The other two first said $50.00, and then, after thinking said something like "no, that's too much, it would be nearer to $10.00 or $20.00." The question seems hard to answer, and perhaps the difficulty is related to the difficulty we observe when we try to get at negative utility. The *S*s who reconsidered seemed to be searching for some rational answer, but also seemed to find it difficult to make a reasonable assessment of the situation. With no other data to go on, I would conclude that the questions are not symmetrical and, therefore, that we may be asking too much of *S* to expect concordance on these judgments.

THE MEASUREMENT OF SUBJECTIVE PROBABILITY

Just as there are errorless techniques for naming the quantity of money by assigning cardinal numbers to quantities of money, so are there errorless techniques for calculating the objective probability of events whose outcomes possess certain symmetrical characteristics or that are well modeled by the calculus of probabilities. For this reason, it is extremely difficult to determine what the subjective probability would be to a person who was asked to assess the likelihood of a die falling with a three uppermost. For those people who understand something about the principles of probability and its application to symmetrical gambling devices, the immediate answer is one-sixth.

Unlike money, where we could assume a subjective counterpart that was not anchored in language, the notion of 'likelihood' is intimately tied to the general theory of probability, especially among our *S*s. This is not to say that they are in fact sophisticated with respect to the theory of probability, but rather that they know they should be sophisticated, and for this reason undoubtedly exhibit biases that would make the determination of a scale of risky likelihood extremely difficult. The problem is to find a collection of events for which the calculation of an objective probability in terms of either *a priori* considerations or relative frequencies is either impossible or extremely difficult. Our questions, then, concerning likelihood, divide themselves into two particular aspects: (1) Do particular uncertain environmental events have a stable subjective probability associated with them? and (2) can we assess these subjective probabilities by the scaling procedures generally employed in psychology?

To answer these questions we performed the following experiments.

A list of 10 uncertain events was prepared. Table II presents the list. For our first experiment, we presented the 10 items pairwise, using all possible pairs of items, to a class consisting of 27 undergraduate women at the University of Pennsylvania, and then to a class of 32 men. They were asked to assert for each pair of items which of the two items was more likely to occur. From these judgments of paired comparisons we were able to assign to the 10 events a number between zero and one reflecting the proportion of times that it was judged more likely than other items in the collection. Column 1 of Table III lists the items; Column 2 shows the proportions for the 10 items obtained from 59 *Ss*.

These same 10 items were then presented individually to each of 25 different undergraduate men of the University of Pennsylvania. Before presenting any of the items to the men, the following instructions were read to them:

> I am going to describe a series of events to you. I would like you to assign a likelihood to these events using the numbers from zero through 100. Zero means you are absolutely certain that the event will not happen, and 100 means you are absolutely certain that the event will happen.

TABLE II

EVENTS TO WHICH THE *Ss* WERE ASKED TO ASSIGN LIKELIHOODS

What is the likelihood that:

(1) It will not rain in Philadelphia for the entire month of April.
(2) Someone could bicycle non-stop from Philadelphia to New York.
(3) The next person that you meet will have had an appendectomy.
(4) You would survive an airplane crash.
(5) The next person you see outside this building will be a blonde.
(6) The National League will win the World Series in 1958.
(7) You would break a bone when falling out of a second story window.
(8) July fourth in Philadelphia will be a hot day.
(9) The next car that you see will have Pennsylvania license plates.
(10) You can break a raw egg with a hammer.

TABLE III

THE VALUES IN SUBJECTIVE UNITS OF THE THREE SCALES OF LIKELIHOOD

Items	Paired Comparisons	100-point rating scale	Magnitude estimation	Average likelihood
(1)	0.04	0.03	0.06	0.04
(2)	0.19	0.36	0.25	0.24
(3)	0.37	0.32	0.25	0.33
(4)	0.32	0.41	0.30	0.34
(5)	0.47	0.32	0.32	0.40
(6)	0.43	0.52	0.50	0.47
(7)	0.65	0.81	0.87	0.74
(8)	0.76	0.88	0.83	0.80
(9)	0.80	0.89	0.99	0.86
(10)	0.95	1.00	1.00	0.97

Thus, the 25 *Ss* were asked to assign rating-scale values to these events using a scale from zero to 100. The average ratings are listed in Column 3 of Table III.

As a final estimate of the likelihoods of the 10 events, we asked 25 additional *Ss*, this time graduate students in psychology and undergraduate psychology majors, to do the following:

> I am going to describe a series of events to you. I would like you to assign a number to each event which will be indicative of its likelihood. You may use any numbers you wish, but I want you to use the smaller numbers to represent the less likely events and the larger numbers to represent the more likely events. I would like you to assign the numbers to the events so that the numbers are proportional to the likelihood. That is to say, if one event is twice as likely as another give it a number twice as large.

The *Ss* in this experiment were asked to make direct magnitude-estimations of the likelihood of the 10 events. Column 4 of Table III contains these magnitude-estimations. The numbers inserted in Column 4 were obtained by taking for each *S* the numbers that he assigned to the events and then multiplying or dividing them by a constant that would make the number assigned to Item (6) equal to 0.50. After this procedure had been carried out, we took the medians of the 25 judgments made by the *Ss* and recorded them in Column 4.

The three scales indicate a fair reliability and an essential linearity between the different kinds of judgments. It seems reasonable, then, to select as the representative likelihood of the 10 events the average of the 3 scale-values. These average scale-values are listed in Column 5.

The results of this experiment suggest that the direct assessment of the likelihood of events with only vague actuarial probabilities is feasible. In addition, the results have another interesting aspect. In an extensive study of the relation between rating-scale and magnitude-estimation techniques, Stevens and Galanter found that for all continua of Class 1 (prothetic), the two scales were nonlinearly related; they also found that on another class of continua (methathetic), rating- and magnitude-scales could be linearly related.[7] The fact that there appears to be a linear relation between the rating- and magnitude-scales in this experiment suggests that the likelihood of events is a Class 2 continuum. This seems quite reasonable in view of the fact that the notion of likelihood has certain symmetrical properties similar to the judgment of the proportionality of colored dots or the judgment of the position of a point on a line. This symmetry may be characteristic of all Class 2 continua. Other experiments on the scaling of perceptual events that contain this principle of proportionality seem to suggest that they too are Class 2 continua.

A final point is the fact that techniques of direct estimation of magnitudes of events for which there are no obvious underlying physical continua is a perfectly feasible procedure. This was suggested in a recent paper by Stevens and the present result lends added credence to the conjecture.[8] There is no need to presume an underlying physical continuum in order to use methods of direct magnitude-estimation.[9]

THE PREDICTION OF CHOICES BETWEEN ALTERNATIVES WITH UNCERTAIN OUTCOMES

Now that we have an ostensible scale for the utility of money and a set of events with known subjective probabilities, it is possible to form a set of alternatives with uncertain outcomes that can be described with two numbers. The question is whether we can predict the pairwise choices that people will make between them. Concretely the question that we ask is: What is the probability that a person will elect the possibility of getting, say, \$64.00 if the next car he sees has a Pennsylvania license plate, versus the possibility of receiving \$217.00 if the next person he meets has had an appendectomy? Since we use 10 different events and there are an indefinite number of sums of money, the number of possible alternatives that we can construct is limitless. We can, however, sample among the events and among certain sums of money and use this sample to determine the predictability of choices.

---

[7] Stevens and Galanter, *op. cit.*, 400.

[8] Stevens, *op. cit.*, 177.

[9] In our initial pilot studies we used a variety of events as candidates for our final

Before we construct the alternatives, it would be better, however, to decide upon the theory of prediction that we propose to test. One suggested basis for choice among uncertain alternatives is the assertion that the probability of choice depends only upon the difference in the subjective expected utilities of the alternatives. The subjective expected utility is computed by taking the product of the utility of the outcome and the likelihood of the event. Since this theory is simple, it seems reasonable to subject it to direct test. If this composition does predict choices, then the theory itself suggests the limits that we should set on the selection or construction of our alternatives. Specifically, we wish to use sets of alternatives whose difference in subjective expected utility is not so large as to make one of the alternatives always or never chosen. That is to say, we wish to construct alternatives whose difference is small enough to give us sufficient uncertainty so that from a sample of people we can estimate the probability of choice between the alternatives.

Using this principle as a criterion, we constructed a set of 16 alternatives, and presented them to each of 107 *S*s in a random order. The *S*s were told:

> I am going to read you a series of paired events. Each event has a monetary value if it occurs and you are to pick the event that will profit you the most. For example, you will be given \$1.00 if the sun rises tomorrow or \$50.00 if a penny turns up heads. If you pick the \$1.00 event, you will get nothing if the sun doesn't rise and the same applies to the \$50.00 event. The idea is to make as much money as possible. To do this, you will have to evaluate the likelihood of each event and the amount of money you can gain by its outcome, so take your time in making these judgments. Consider that each event takes place simultaneously; your task then is to make as much money as possible. Are there any questions?

From these pairwise comparisons, we computed the proportion of times that one side of each of the 16 alternatives was selected. This proportion is plotted against the algebraic difference in the subjective expected utility of the alternative pairs in Fig. 2.

---

10 items. Some of the items that we finally used were embedded, therefore, in different context in these early experiments. Interestingly, these items seemed to have their likelihoods unchanged by changes in context. It is also noteworthy that some items of extremely small actuarial value (no rain for the month of April) received likelihoods as high as they did.

DISCUSSION

The results of this series of experiments reveal that it is possible to assess directly the utility of money; further, it is possible to scale the likelihood of uncertain events, and finally, it is possible to use these psychological scale-values to predict the probability of pairwise choice among alternatives with uncertain outcomes. In addition, the results support, by revealing a relatively simple function underlying the prediction of choices, the hypothesis that the subjective expected utility of an uncertain alternative is one major determiner of its probability of selection.

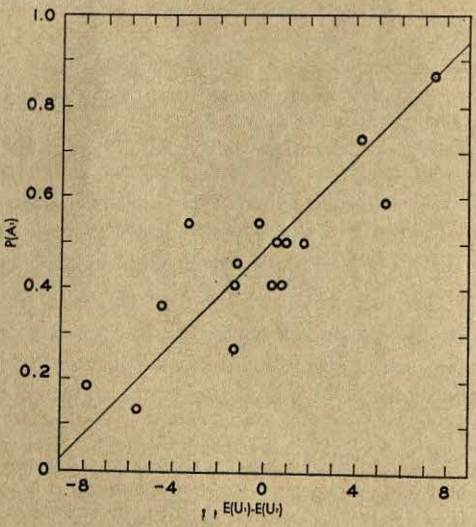


FIG. 2. THE RELATION BETWEEN THE DIFFERENCES IN THE SUBJECTIVE EXPECTED UTILITY OF ALTERNATIVE 1 [$E(U_1)$] AND THE SUBJECTIVE EXPECTED UTILITY OF ALTERNATIVE 2 [$E(U_2)$] PLOTTED AGAINST THE PROPORTION OF SUBJECTS CHOOSING ALTERNATIVE 1 [$P(A_1)$]

(The straight line is merely a visual aid.)

The results also lead to other conclusions. Specifically, these results lend a certain validity to the subjective scale of value for money that we constructed originally. After all, it might very well have been the case that the psychological value of money was a logarithmic function of money, or that the psychological value of money was a linear function of money. In either of these two cases we would have seen a lawful relation between the subjective expected hypothesis of utility and the probabilities of choice. In both cases, however, the relation would have been more complex than the relation that we observe.

To examine this issue, we computed the expected values of the alternatives that the *S*s were given on the assumption that the utility of money was a logarithmic function of money with parameters approximating those we used for the power-function. We also computed the expected value of the alternatives assuming that the value of money was linear with money. We then replotted Fig. 2, using these new differences in expected value and assumed that all three functions were linear. As an index of quality, we computed the rank-order correlation between probability of choice and these differences in the three expected utilities. The scattergrams of the two new scales of utility are plotted in Fig. 3. The correlation between the difference in expected utility as a power-function of money and choice is +0.88. With the expected utility equal to the logarithm of money, the correlation is +0.71; while against the differences based on a linear relation the correlation is +0.41.

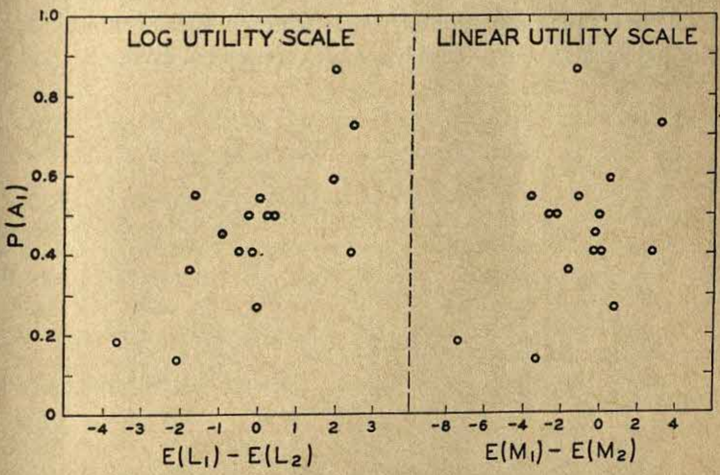


FIG. 3. SCATTER DIAGRAMS OF TWO SCALES OF UTILITY IN THE LOG $[E(L)]$ AND LINEAR $[E(M)]$ SUBJECTIVE EXPECTED UTILITY-DIFFERENCES PLOTTED AGAINST THE PROPORTION OF SS CHOOSING ALTERNATIVE 1

The question of the exact value or the distribution of values of the exponent of the utility-function must await further research. It seems intuitively obvious that the exponent will be less than one, but the numerical value for the typical *S* is hard to determine. A question we plan to explore is whether we can predict with greater precision by estimating this parameter for each *S* and using it to predict large numbers of his choices.

A final question about the value of money concerns the value of negative quantities of wealth. It seems likely that the decrease in value as loss increases will grow as some power greater than one. An interesting study would be to determine how nearly this power approximates the root of the positive value function. A result of this kind means, of course, that we are doomed to a world where equal monetary increment of gain and loss leave us in a position poorer than our starting point.

In sum, the present study is an experimental demonstration of how far it is possible to go using direct scaling-techniques. We have obtained scale-quantities for the value of money, and a scale for the likelihood of events, and using these scale-values we have been able to predict within reasonable limits the proportion of people who will make certain kinds of complex judgments about ostensibly complex conceptual events.

EXPERIMENTAL REDUCTION OF AUTOKINETIC MOVEMENT

By JOSEPH R. ROYCE, University of Alberta, WILLIAM R. STAYTON, University of Redlands, and ROBERT G. KINKADE, Ohio State University

In spite of the fact that the autokinetic phenomenon has been experimentally studied since the dawn of modern psychology, there is little precise knowledge concerning the nature of this phenomenon.[1] Perhaps the major reason for this is that the bulk of the research in this area has been qualitative. Until the present decade, the primary method of recording the extent of movement was by verbal report. Similarly, little systematic effort has been devoted to determining quantitatively the exact conditions which result in 'freezing' or 'reduction' of this illusion.

The studies of Graybiel and Clark include observations on the reduction-effect of multiple light patterns on autokinetic movement.[2] This study involved one of the few efforts to quantify responses; namely, the procedure of depressing a key so connected to a timer that the length of time movement was perceived could be recorded. The effects of a diverse array of multiple light patterns and stimulus-figures (*e.g.* large A-light, semi-circle, rectangle, cross, etc.) have been reported in the researches of such authors as Adams, Crutchfield and Edwards, Karwoski and associates, and Luchins.[3] Of these, perhaps the study most closely related to our

---

* Received for publication August 31, 1959. Based in part on work done under ARDC Project 7707 at the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. This research was also supported by the National Institute of Mental Health.

[1] The first report of autokinetic movement was in 1799 by the astronomer, Von Humboldt. Sweizer was the first to point out that it was an illusion (1858). Charpentier was the first to move the observations into the laboratory situation (1886), and Aubert gave it its name (1887). For a brief summary of this early research, see H. F. Adams, Autokinetic sensations, *Psychol. Monogr.* 14, 1912, 1-45, and J. P. Guilford and K. M. Dallenbach, A study of autokinetic sensation, this JOURNAL, 40, 1928, 83-91.

[2] Ashton Graybiel and Brant Clark, The autokinetic illusion and its significance in night flying, *J. Aviation Med.*, 10, 1945, 111-151.

[3] Adams, *op. cit.*, 1-45; R. S. Crutchfield and Ward Edwards, The effect of a fixated figure on autokinetic movement, *J. exp. Psychol.*, 39, 1949, 561-568; Differential reduction of autokinetic movement by a fixated figure, *ibid.*, 41, 1951, 25-31; Ward Edwards, Two and three dimensional autokinetic movement as a function of size and brightness of stimuli, *ibid.*, 48, 1954, 391-398; Autokinetic movement of very large stimuli, *ibid.*, 48, 1954, 493-495; T. F. Karwoski, Harold Redner, and H. O. Wood, Autokinetic movement of large stimuli, *J. gen. Psychol.*, 39, 1948, 29-37; A. S. Luchins, The relation of size of light to autokinetic effect, *J. Psychol.*, 38, 1954, 439-452; The autokinetic effect and gradation of illumination of the visual field, *J. gen. Psychol.*, 54, 1954, 29-37.

work is one of Luchins, who has conducted the most thorough investigation on the reduction of movement due to the increase of light-intensity in the visual field.[4]

The purpose of this research is to provide a more adequate quantitative procedure for measuring autokinetic movement, and further, to investigate the effect of physical frames of reference on the autokinetic phenomenon. As a working definition, we will think of physical frame of reference as stimuli in the visual field in addition to the autokinetic light. The frame of reference then, can consist of a second pinpoint of light, a configuration of lights, a concentric circle, a straight line, etc. The experiments which follow, especially Experiments I and II, are exploratory.

EXPERIMENT I

The purpose of this experiment was to determine the effect of various patterns of lights on autokinetic movement.

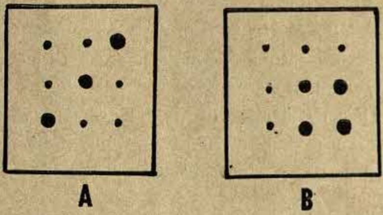


FIG. 1. TWO SAMPLE CONFIGURATIONS
The lights were numbered by row from left to right and top down; thus Sample A is designated as 3, 5, 7, and Sample B as 5, 6, 8, 9.

Apparatus. The apparatus consisted of a square 3 × 3 matrix of nine small lights, distributed at 9-in. intervals. The lights were 1 mm. in diameter. The experimental room was light-proofed, 7.5 ft. wide and 12.5 ft. long. *S* was seated in a chair 6 ft. from the apparatus and directly in front of the stimulus-field. Responses were recorded by *E* with the aid of a shadow box. There were 10 *Ss*, 9 students and one professor of psychology. Four of the students knew nothing before the experiment about the autokinetic phenomenon.

Procedure. The *Ss* were brought into the experimental room one at a time. While dark adapting for 5 min. they were told that they should report any movement or non-movement of the single light or pattern of lights they were about to view. In addition to the autokinetic light, nine different configurations shown in Fig. 1 and Table I, involving from two to nine lights, were presented. In some cases, all 10 of the *Ss* viewed a given light-pattern; in other cases, only half the *Ss* were involved. The configurations were given in order at one sitting, beginning with one light and ending with nine. The exposure-time for a given observation was 3 min. followed by an immediate oral report regarding movement. Since a given *S* was exposed to only 7 of the 10 experimental conditions, the total time for each *S* was around 30 min. In every case except two patterns the *Ss* were instructed to fixate at the center of the configuration. In these two cases the *Ss* were instructed to fixate the light which was at the right angle of these two triangular configurations.

---

[4] Luchins, *op. cit.*, *J. gen. Psychol.*, 54, 1954, 29-37.

Results. The results of Experiment I are summarized in Table I. Of the *Ss*, 90% observed movement of the pinpoint of light alone and only 20% reported movement when 9 lights were presented. In general, fewer *Ss* reported movement as more lights were exposed (Spearman rho = −0.65). The exception to this occurred in the case of the configuration involving five lights (No. 1, 3, 5, 7, 9). The distance between corner lights was doubled (see Fig. 1 and Table I), providing a much less tight stimulus-configuration for *S*. This might account for the increase in observation of movement. Introspective reports indicated that movement of four lights or more (the last three patterns of Table I) provided a shaky movement of the total configuration. For example, when all nine lights were presented, *S* described the movement as "fluid."

TABLE I

PERCENTAGE OF *Ss* OBSERVING MOVEMENT FOR SPECIFIED LIGHT-CONFIGURATIONS

Configuration of lights*	No. of lights in configuration	% *Ss* reporting movement	No. of *Ss*
A-light 5	1	90	10
5, 6	2	80	10
2, 5	2	60	5
1, 5	2	80	5
5, 6, 8	3	60	5
3, 5, 7	3	80	5
2, 4, 5	3	80	5
5, 6, 8, 9	4	30	10
1, 3, 5, 7, 9	5	60	10
All	9	20	5

* See Fig. 1 for explanation of configurations.

EXPERIMENT II

The purpose of Experiment II was to determine the effect of varying the radius and the order of presentation of a ¼-in. wide concentric circle on autokinetic movement. Since a concentric circle is a continuous and symmetrical configuration, it would appear to be the ideal physical frame of reference in terms of the purposes of this investigation.

Apparatus. S was seated 7 ft. from a display consisting of a light at the center surrounded by a concentric circle of light. The circle could be varied in size. The width of its circumference was held constant at ¼ in. while the radius varied among three different values, 1.5, 3, and 4.5 in. Light intensity was also held constant during this series. The *Ss* used for this experiment were 18 college students who were completely naïve about the autokinetic phenomenon.

Procedure. The *S*s were randomly distributed among three different groups, according to radius of concentric circles (see Table II). There were two parts to the experiment, which will be referred to as Set 1 and Set 2. The first part consisted of showing the pinpoint of light to *S* *until* he had observed movement for 1 min., followed by a 3-min. presentation of a circle of light along with the A-light. The second part of the experiment involved a reversal of the first part. The circle was shown alone for 3 min., followed by a 1-min. presentation of the concentric circle together with the A-light. All *S*s were brought into the experimental room one at a time and dark-adapted for 5 min. before Set 1 and 2 min. before Set 2. They were instructed to report whether or not they saw movement of the A-light before and after presentation of the concentric band of light.

Results. One *S* was rejected because of myopia. The experiment made her dizzy and her judgments were very erratic; it was felt, therefore, that her responses were invalid. The results of the 17 *S*s whose data were analyzed are summarized in Table II for Set 1 and Set 2 respectively. It is immediately apparent that the over-all trend of the data is the same for both sets. In other words, the perceived movement is essentially the same, regardless of radius and whether the autokinetic light is presented first and then surrounded by a concentric circle of light, or whether the concentric circle is presented first followed by exposing the autokinetic light in its center.

TABLE II

NUMBER OF *S*S PERCEIVING AUTOKINETIC MOVEMENT UNDER CONDITIONS OF SET I AND SET II

Type of movement	Radius of circle 1.5 in.		3.0 in.		4.5 in.	
	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2
Independent	2	2	1	0	1	1
Coördinate	2	2	4	5	1	2
None	1	1	2	2	3	2
Total *N*	5	5	7	7	5	5

While more *S*s failed to see movement at a radius of 4.5 in., it would be misleading to draw firm conclusions on the basis of these preliminary observations. The important point is that at all radii there was reduced movement, regardless of set. Most *S*s continued to see movement of some type (65% for Set 1 and 71% for Set 2) after the introduction of the concentric circle. The greater proportion of the movement observed involved perception of the total configuration, that is, movement of both the autokinetic light and the concentric circle, with the light 'frozen' inside the circle. This has been designated in the table as "coördinate movement." Because of the lack of independent movement of the autokinetic light in the case of coördinate movement, it could be argued that the autokinetic phenomenon has, in fact, been immobilized by the concentric circle. This point is brought out more clearly when we combine all cases of coördinate perception (*i.e.* coördinate movement and no movement), revealing that 76% and 82% of the *Ss* regarded the autokinetic light as essentially frozen. This means that only 24% and 18% of the *Ss* reported movement of the autokinetic light independent of movement of the concentric circle. Introspective reports indicate that this independent movement is rather elusive. When the concentric circle was turned on, movement of the autokinetic light was momentarily stayed for everyone. Those who reported independent movement, however, felt that the autokinetic light had a very slight 'fluid' movement. As will be indicated in Experiment III, this movement is confirmed under a more accurate recording procedure, but it occurs infrequently (see Table III).

TABLE III

THE EXTENT OF AUTOKINETIC MOVEMENT AS A FUNCTION OF RADIUS OF CONCENTRIC CIRCLE

(N=5 for all groups except for the radii of 2 in. and 13 in. where N=4. Movement is measured in centimeters.)

Radius of concentric circle[*]	Base movement[†]	Movement after circle introduced			Percentage reduction of total movement[†]
		total	coördinate	independent	
1	11.10	13.40	13.40	.00	−23.26
2	27.12	2.75	2.00	.75	77.78
3	20.00	5.00	4.50	.50	1.60
4	17.70	12.40	12.40	.00	30.40
5	25.60	15.00	15.00	.00	48.60
6	13.10	6.30	5.10	1.20	48.40
7	19.80	12.10	3.30	8.80	43.40
8	23.88	9.82	8.22	1.60	43.60
9	15.20	8.40	5.30	3.10	48.80
10	21.90	10.86	6.00	4.86	45.60
11	15.00	2.80	.80	2.00	86.00
12	12.80	1.82	1.30	.52	84.20
13	10.02	1.05	.50	.55	88.50
Total		101.70	77.82	23.88	

[*] Base movement refers to the extent of movement of the A-light alone, before the concentric circle was placed in the visual field.

[†] These figures are averages of the percentage reduction of individual cases, not percentage reduction averages. Minus figure represents an increase in total movement.

EXPERIMENT III

Experiment III is an extension of Experiment II under more adequate conditions of presenting stimuli and recording responses. We were particularly interested in a more thorough check on the effect of the radius of concentric circle on autokinetic movement. We also made preliminary observations of the effect of intensity of the autokinetic light relative to intensity of concentric circle light.

Subjects. There was a total of 103 *Ss*, 58 unclassified basic trainees of the Air Force located at Lackland Air Force Base, San Antonio, Texas, and 45 college students from the University of Redlands.[5] In this experiment *S* participated in only one observation to provide control over the possible effects of one observation on a subsequent one. The exact number of *S*s per experimental condition is given in the tables summarizing the results. There were either 4 or 5 *S*s per condition.

Apparatus. How can we translate subjective movement into measurable behavior?

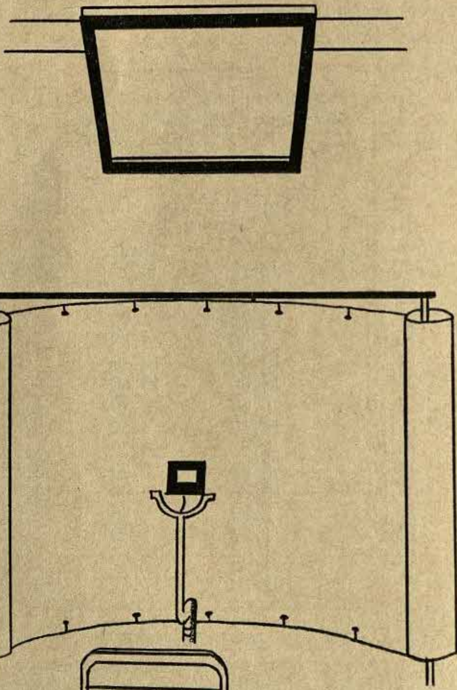


FIG. 2. THE RECORDING UNIT

The most elaborate answers to this question are the forms of tracking devised by Guilford and Dallenbach, Bridges and Bitterman, and by Conklin.[6] None, for various reasons, suffices, hence we devised another method of measurement which is more flexible from the point of view of the stimulus-field, and which is also objective. This method is fundamentally an elaboration of simple tracing and can best be explained by reference to Figs. 2 and 3.

---

[5] The data gathered on the 45 college students for Experiment III were obtained by Stanley Rule.

[6] Guilford and Dallenbach, *op. cit.,* 86 ff.; C. C. Bridges and M. E. Bitterman, The measurement of autokinetic movement, this JOURNAL, 67, 1954, 525-529; J. D. Conklin, Quantitative determination of autokinetic movement: rate, direction, extent, and latency, *Northwest Science,* 30, 1956, 61-69.

(1) Mirror-system. Stimuli are presented to *S* from the rear via two front-silvered mirrors. A large (24 × 24 in.) adjustable mirror is mounted on the ceiling, and a small adjustable mirror is mounted on the chin-rest. The system of mirrors allows complete freedom in tracing movements without obstruction from the stimulus-unit.

(2) Adjustable chin-rest. *S* sits in a chair in front of the apparatus in such a way that he is comfortable with respect to the position of his chin in the chin-rest.

(3) Tracing board. The record of movement is traced with a pencil on paper covering a large board in front of *S*. The paper is wound on two 36-in. spindles located on either side of the apparatus. The paper is unwound from the spindle on the right, run through ball bearing rollers which hold the paper firmly in position on the tracing board, and wound onto the spindle on the left. The backing board is a 6 ft. wide, ⅛ in. thick piece of plywood which has been bowed in such a way as to describe a left-to-right arc the length of a man's arm. The distance between the two spindles is 4 ft., providing approximately a 2-ft. curve in the concave writing surface. The height of the plywood is 4 ft.

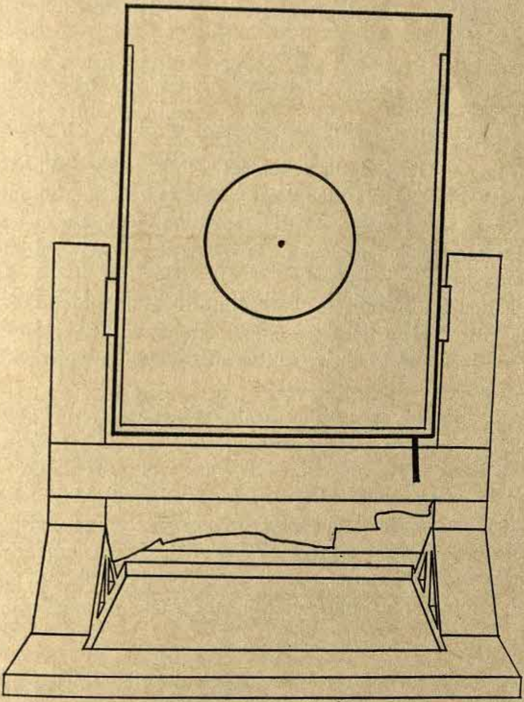


FIG. 3. THE STIMULUS-UNIT

(4) Measurements. Rate, direction, and extent of movement are measured directly from the trace. The autokinetic stimulus is presented for a standard period of time, and the rate of movement is simply the distance divided by this time. Latency is

measured by an arrangement which is not shown in Fig. 2. When the stimulus-light is turned on, a timer is started. *S* operates a hand switch as soon as the stimulus-light appears to move and he simultaneously tracks the movement with the other hand.

The stimulus-unit (Fig. 3) consists of a ceramic glass panel lighted from behind by 60 six-volt A.C. light bulbs (Mazda 47). The panel is 30 in. wide and 36 in. high. Cardboard cut-outs are mounted in front of this field by sliding them into the grooved edges of the display-panel. The autokinetic light, 1 mm. in diameter, controlled by a separate circuit, is mounted at the center of the stimulus-unit. Light intensity was controlled by a Variac transformer in the electrical supply to the lights.

Procedure: (1) Concentric circles. Groups of 4 to 5 *Ss* were presented with an autokinetic light surrounded by a ¼-in. concentric circle with 13 different radii ranging from 1 to 13 in. Light intensity was held constant, *i.e.* the concentric circle with a Variac reading at 22 and the autokinetic light with a Variac reading at 55. The intensity-value of 22 was the least value capable of holding *S*'s attention. We will refer to this intensity as the "attention-threshold." The intensity of 55 for the stimulus for which movement was reported was selected simply because it was fairly bright and readily perceived.

After dark-adaptation for 20 min. in the waiting room adjoining the experimental dark room, *S* was instructed to track the autokinetic movement under the condition of the A-light alone followed by the condition of the A-light surrounded by a concentric circle.

(2) Light-intensity. In the experiments with light-intensity, the concentric circle was held constant (a ¼ in. wide circle with a radius of 7 in.) while the intensities of both autokinetic light and concentric circle light were varied. Two series of experiments, A and B, were conducted. In Series A, the autokinetic light was held constant at the attention-threshold and the concentric circle was varied through four intensities.[7] Groups of 5 *Ss* were exposed to one of the four conditions. In Series B, the concentric circle was held constant at the attention-threshold and the autokinetic light was varied through four intensities in the same manner as described for Series A.

The instructions to all the *Ss* were the same as described in the concentric circle experiments noted above.

Results. The effect of varying radius on the extent of autokinetic movement is summarized in the data presented in Table III. In general, the results clearly show that a reduction of movement occurs when the autokinetic light is surrounded by a ¼-in. wide concentric circle of light, for nearly all radii of the circle. This point is brought out most clearly in the last column. Here we see that the percentage of reduction in movement is considerable, ranging from −23.3% to +88.5% with a median value of

---

[7] The lowest intensity was a value approaching the lower threshold; the maximum was a degree of illumination that approached the amount of light required to see other objects in the experimental room.

48.4%.[8] The only observation which does not result in a reduction of movement is the case where the radius is 1 in. In this case the extent of movement increased by 23.26%.

To check whether the reduction in movement was significantly different for different radii, an analysis of covariance was carried out. This involved working directly with the raw scores on base movement and total movement for each of the 63 *S*s rather than with the scores of percentage-reduction given in the last column of Table III. The *F*-values obtained were as follows: base movement, 0.555; total movement, 1.163; both involving 12 × 50 df, and the covariance of base and total movement, 1.178 with 12 and 49 df. None of these values is statistically significant, indicating that the *S*s were randomly assigned to the 13 groups, that differences in the radius of the concentric does not have significant effect on reduction of movement, and that the ratio of total movement to base movement also fail to show a differential reduction as a function of the radius.

We conclude, therefore, that concentric circles with radii ranging from 1 in. to 13 in. are essentially equal in their effect on the extent of autokinetic movement. Inspection of the data in Table III suggests that the larger radii might be more effective in reducing movement. For example, the rank-order-correlation (rho) between radius of the circle and percentage-reduction is 0.69. This is true because the total or coördinate movement decreases. There is so little independent movement perceived at any radius that no noticeable effect can be shown.

As in Experiment II, by far the greater proportion of *S*s who saw movement observed coördinate (43/52 or 83%) as opposed to independent (13/52 or 25%) movement.[9] There were only 6 of the 52 (11.5%) who saw both. The data on extent of movement in Table III also reveal the predominance of coördinate 77.82/101.70 or 76.51%) over independent (23.88/101.70 or 23.4%) movement. If we argue, as we did previously, that coördinate movement means that the autokinetic light is actually 'frozen' inside the circle, and regard independent movement as the only true A-movement, the influence of the concentric circle is even more marked in practically every case.

In Table IV, we have summarized the data concerned with the effect of light-intensity upon autokinetic movement. It is doubtful whether any of the differences among values in this table is statistically significant. The only consistent trend in these data relates to coördinate autokinetic movement as dependent upon the intensity of the surrounding concentric circle. The data suggest that the greater the light intensity of the concentric circle, the less the extent of coördinate movement. Furthermore, the proportionate amount of independent movement in Table IV is greater than that observed in Table III (42% when the concentric circle was varied and 35% when the A-light was varied, as contrasted with 23.4% in Table III). It is presumed that the increase in independent movement under the conditions reported in Table IV is in some way attributable to the contrast in the intensity of the autokinetic spot and the concentric circle.

---

[8] Percentage-reduction of movement was computed separately for each *S* by subtracting total movement from base movement, and dividing by the appropriate base movement.

[9] Eleven of the 63 *S*s did not see any movement, either coördinate or independent, after introduction of the concentric circle. This accounts for the *N* of 52 rather than 63 in computing these proportions.

TABLE IV

EXTENT OF AUTOKINETIC MOVEMENT AS A FUNCTION OF THE LIGHT INTENSITY OF THE CENTRAL DOT AND OF THE CONCENTRIC CIRCLE.

(*N* is 5 for all observations.)

	Type of movement			
Intensity	coördinate		independent	
	A-light	circle	A-light	circle
Minimal level	3.36	4.90	.00	2.74
Intermediate level 1	.58	3.06	5.76	.80
Intermediate level 2	5.02	2.12	.88	2.90
Maximal level	4.48	.20	.90	1.00
Totals	13.44	10.28	7.54	7.44

DISCUSSION

Perhaps the most widely accepted explanation of autokinesis is that *S* has no anchor for his observation of the spot of light in his visual field. He has no frame of reference, and therefore, the autokinetic light is free to wander. Sherif uses this explanation as the starting point for his further experiments on personal and social factors that influence perception.[10] Little has been done in a systematic way to determine experimentally what are the effects of various physical frames of reference.

Two extremes of the continuum from less to greater strength of the frame of reference were explored; namely, the presence of another pinpoint of light as a very vague and discontinuous frame of reference, and a concentric circle of light as a very clear and continuous frame of reference. The results confirm the hypothesis that autokinetic movement will be reduced when the autokinetic light is provided with a frame of reference. Furthermore, the more continuous the frame of reference, the greater the reduction-effect. This result is consistent with a number of earlier observations. For example in one of Edwards' studies, less independent movement is reported for an A-light surrounded by a matrix of A-lights as contrasted with an A-light surrounded by a rectangle (*i.e.* non-filled matrix) of A-lights.[11]

---

[10] Muzafer Sherif, A study of some social factors in perception, *Archiv. Psychol.*, 27, 1935, (No. 187), 1-60; and *The Psychology of Social Norms*, 1936.

SUMMARY

In Experiment I, we determined the effect of various patterns of lighted dots on autokinetic movement. In Experiments II and III, we determined the effects on autokinetic movement of (a) varying the radius of a ¼-in. wide concentric circle of light about a central dot, and (b) varying the intensities of both the centered dot and the concentric circle, holding the radius of the concentric circle constant.

The use of a physical frame of reference in this study resulted in a reduction or freezing of autokinetic movement. In particular, the extent of perceived movement decreases: (1) as the number of lights in a multiple-light configuration increases; (2) in the presence of a ¼-inch wide concentric band of light, regardless of radius; and (3) as the light-intensity of the concentric circle increases.

---

[11] Edwards, *op. cit., J. exp. Psychol.,* 48, 1954, 493-495.

THE INFLUENCE OF HUE, LIGHTNESS, AND SATURATION ON APPARENT WARMTH AND WEIGHT

By BENJAMIN WRIGHT, University of Chicago

The belief that color can impart such effects as apparent warmth and weight is widespread and ancient. It is customary, for example, to speak of the warm yellows, oranges, and reds in contrast to the cool violets, blues and greens. There have been a number of experimental studies aimed at determining whether colors can in fact impart such effects. Unfortunately, most of them have been unsatisfying in two fundamental respects.

The major shortcoming has been the failure to deal explicitly and successfully with the three basic perceptual dimensions of color: hue, brightness (called "lightness" when reference is made to surface color), and saturation. As Arnheim says:

> The expression of color in general and its temperature in particular are influenced not only by hue but also by brightness and saturation. Therefore the expressive values of hues can be compared only when the two other factors are kept constant. For example . . . there is some indication that a high degree of brightness tends to make a color cold and a low degree of brightness tends to make a color warm. Thus in order to be sure, for example, that a pure red is warmer than a pure yellow we would have to compare them at equal brightness.[1]

Most previous studies have employed so few colors as to make such a task impossible. Of the earlier studies, only those of Monroe and of Ross introduced all three color-dimensions into the structure of their observations.[2] Neither of them, however, succeeded in untangling the separate effects of each dimension in their analyses. Monroe noted the confounding of hue and lightness in her design but did nothing about it. Ross overlooked the high correlation between hue and brightness among his colors and arrived at conclusions inconsistent both with other research and with his own data.

A second difficulty has been the confusion of different modes of perception. Three previous studies report inconsistencies between strictly visual and visual-tactual or visual-kinesthetic color-effects. These inconsistencies suggest color-weight and color-warmth illusions similar to the size-weight illusion.

For example, when Mogensen and English asked their *Ss* to judge the tactual warmth of equally warm thermostats wrapped in differently colored paper, they found, contrary to their expectations, that the green and blue thermostats were

---

* Received for publication November 10, 1960. The collection of the data for this study was sponsored by Social Research, Inc., Chicago, Illinois, and by M. F. and P. F. Reemtsma, Hamburg, Germany. The analysis of the data was sponsored by the Social Science Research Committee of the University of Chicago.

[1] Rudolf Arnheim, *Art and Visual Perception,* 1954, 277-278.

[2] Marion Monroe, The apparent weight of color and correlated phenomena, this JOURNAL, 36, 1925, 192-206; R. T. Ross, Studies in the psychology of the theater, *Psychol. Record,* 2, 1938, 127-190.

judged slightly warmer than the red, yellow, and orange ones![3] When De Camp and Taylor asked *Ss* to judge pairs of equally weighted but differently colored cubes by 'hefting' them, they found that light colored cubes tended to be judged heavier than dark ones.[4] Taylor noted that "there were some observers for whom the colors in order of apparent heaviness ranged from darkest to lightest in the 'looks' procedure and approximated a reverse arrangement in the 'hefting' procedure,"[5] and, that "there were several persons whose spontaneous remarks indicated that they had been consciously trying to disregard the colors in making their judgments."[6] Warden and Flynn investigated the effect of color and serial arrangement on the apparent size and weight of cartons which were identical in size and shape. The judgments of the *Ss*, made visually, revealed that the color-size illusion did not depend upon color but upon the serial arrangement of the cartons; and that the color-weight illusion occurs independently of the arrangement of the cartons and appears to be due to color.[7]

Such ambiguous results as these suggest that color can lead to the systematic misjudgment of concomitant perceptions due to an over-compensation for the color-effect. When tactual or kinesthetic judgments are either inadvertantly provoked by observational conditions or explicitly used to represent the color effect, the resulting data can become confusing.

This paper has two purposes: (1) to bring together and, when necessary, to reinterpret previous visual studies; and (2) to report the results of a study which bears on some of the shortcomings of previous work. In brief: (a) the present study covered 45 different colors, as well as black, white, and gray. (b) Observations were focused strictly on visual judgments so that the influence of color-warmth and color-weight illusions was minimal. (c) In analyzing the results, the separate effects of hue, lightness, and saturation were examined by means of multiple regression. After a brief report of this study, warmth and weight will be discussed in turn.

MATERIALS AND PROCEDURE

Subjects. The *Ss* were middle and lower class men and women living in urban West Germany in 1957. Each *S* judged a single 3-in. square of matte, surface color presented in daylight against a neutral background. Judgments were expressed on an Osgood semantic differential by circling one of seven positions between pairs of polar adjectives like these:[8]

KALT ... . WARM
LEICHT ... . SCHWER

---

[3] M. F. Mogensen and H. B. English, The apparent warmth of colors, this JOURNAL, 37, 1926, 427-428.

[4] J. E. DeCamp, The influence of color on apparent weight, *J. exp. Psychol.*, 2, 1917, 347-370; C. D. Taylor, Visual perception versus visual plus kinesthetic perception in judging colored weights, *J. gen. Psychol.*, 4, 1930, 229-245.

[5] Taylor, *op. cit.*, 241.

[6] *Ibid.*, 240.

[7] G. J. Warden and E. L. Flynn, The effect of color on apparent size and weight, this JOURNAL, 37, 1926, 398 ff.

[8] C. E. Osgood, G. J. Suci, P. H. Tannenbaum, *The Measurement of Meaning*, 1957, 76-85.

For the purposes of scoring, the seven rating positions between each adjective-pair were weighted uniformly one through seven.

Individual interviews were obtained on a door-to-door basis within the structure of an area probability sampling plan for urban West Germany. Each *S* rated just one of the 45 colors, black, gray, or white on one of two different 24 adjective-pair rating-forms. Sample size per rating-form varied from 20 to 70, with an arithmetic mean of 36. Altogether 926 men and 2643 women participated in these ratings. Age ranged from 16 to 65 yr., with 70% being between 25 and 59 yr. The age and sex of the *S*s for the different colors were comparable. Since the evidence for sex- and age-differences in ratings among the *S*s was secondary, the data are reported for the group as a whole.

In Munsell color coördinates, hue had a mean of 31 and a standard deviation of 23; lightness had a mean of 5 and a standard deviation of 2.1; and saturation had a mean of 9 and a standard deviation of 3.8. The color-dimensions were relatively uncorrelated with one another over the 45 colors. Hue had a correlation of -0.11 with lightness and of -0.29 with saturation. Lightness had a correlation of $+0.13$ with saturation. Thus the color solid was well filled by the 45 colors and there was little artificial dependence among color dimensions.

RESULTS

Two of the 48 adjectival pairs, kalt-warm and leicht-schwer, bear on apparent warmth and weight. The mean square estimates from average ratings on these adjectival-pairs for among colors (MS_{AC}), linear regression (MS_{HLS}), and residual from linear regression (MS_{res}) are given in Table I. The partial regression coefficients of average ratings on hue, lightness, and saturation, as well as the average differences between ratings of white and black, are given in Table II.

The mean squares in Table I are on a per *S* basis. To evaluate them it is necessary to determine the magnitude of the variance of the per *S* rating to be expected within colors and among *S*s. When a seven-point scale is weighted uniformly from one to seven, the maximal variance per rating is nine. This occurs when only the extremes of the scale are used and these equally often. When all seven positions on the scale are used with equal frequency, then the variance per ratings is four. In several hundred spot-checks among the 2304 samples obtained in West Germany, as well as from similar studies in the United States, the within-colors adjective-pair rating mean-square estimates of the *S*s fluctuated between the limits two to five with about 80% falling between three and four.

The magnitudes of MS_{AC} in Table II can be used as a measure of the degree of consistency with which the *S*s used each adjective-pair to rate colors. Similarly, the magnitudes of MS_{HLS} can be used as measures of the amount of this consistency which can be determined by a linear function of hue, lightness, and saturation. Finally the magnitudes of MS_{res}

bear on the question of whether a linear function is sufficient to account for the consistent part of the variation in average ratings.

Reducing the estimate of residual from linear regression-variation in Table I by the estimated within-color, *i.e.* individual, variation suggests that for apparent warmth linear terms explain at least half of the consistent variation while for apparent weight they explain about two thirds. It may be possible to estimate additional nonlinear terms in the functional relation between hue, lightness, and saturation and apparent warmth and weight, but these data suggest that the linear terms account for the major portion of the consistent variation.

With respect to statistical significance, the 0.0001 percentiles of the null hypothesis variance-ratios with 3 and 41 and with 41 and several hundred degrees of freedom are about 9 and 2 respectively. Assuming a MS_{WC} of the order of 4, all of the mean squares in Table I are highly improbable under the null hypothesis.

TABLE I

DEPENDENCE ON COLOR OF THE RATINGS FOR AVERAGE WARMTH AND WEIGHT

Adjective-pairs	Mean squares			Multiple correlation
	Among colors	Linear regression		
	MS_{AC}*	MS_{HLS}†	MS_{res}‡	R_{HLS}§
kalt-warm	19.8	105	13.5	.59
leicht-schwer	39.3	370	14.7	.81

* Mean square due to ratings among colors with 44 *df*.

† Mean square due to linear dependence on the three color dimensions (hue, lightness, saturation) with 3 *df*.

‡ Residual mean square deviation from regression with 41 *df*.

§ Multiple correlation of adjective-pair rating with hue, lightness, and saturation with 3 and 41 *df*.

The partial regression coefficients given in Table II estimate the degree of linear relation between each of the three color dimensions and the judgments represented by each adjective-pair. These coefficients are computed so that each one gives the relation between a given judgment and a particular color dimension free from the effects of the other two dimensions of color. Thus in Table II, one can examine the independent effect of each color dimension on apparent warmth and weight.

The differences in average white and black ratings given in Table II are expressed in the form of estimates of the Munsell per unit effect of lightness in order to make them comparable to the regression coefficients on lightness computed over the 45 colors.

The results given in Table II will be discussed as warmth and weight are each taken up.

DISCUSSION AND CONCLUSIONS

Warmth. With respect to visual judgments of projected colors, Stefănescu-Goangă, Lewinski, and Ross found that reds were rated warmer than blues.[9] Only Ross, however, kept track of all three color dimensions and his report requires interpretation. He concluded that "temperature is related to the brightness and saturation of color" and reported correlations which imply that increasing ratings of warmth correspond both to increasing saturation and to increasing brightness.[10] These correlations, however, are not free from the effects of hue. Brightness and hue correlate about −0.8 among the colors Ross used.

TABLE II

LINEAR RELATION BETWEEN WARMTH AND WEIGHT RATINGS AND HUE, LIGHTNESS, AND SATURATION

Adjective-pairs	Regression-coefficients on color dimensions*						White compared to black
	hue		lightness		saturation		
	$b_{H.LS}$	s_b	$b_{L.HS}$	s_b	$b_{S.HL}$	s_b	$W\text{-}B/9$
kalt-warm	*.169*	.046	.090	.050	−.041	.028	.14
leicht-schwer	.033	.048	*.436*	.052	.042	.029	.46

* Explanation of symbols used in table: $b_{H.LS}$, $b_{S.HL}$, $b_{L.HS}$: partial regression coefficients of average adjective-pair ratings on one color dimension after the variation due to the other two has been removed. s_b; standard error of partial regression-coefficients based on the residual from regression mean square with 41 *df*. $W\text{-}B/9$: average rating for white minus average rating for black divided by the difference in lightness of nine units between the white and black color samples to give an estimate comparable to the regression coefficient. The standard error of this estimate, based on within-color variance with 98 *df*, is 0.05.

A careful reëxamination of Ross's data in the red and in the blue regions separately shows: (1) that there is no evidence for a relation between warmth and either brightness or saturation in the blue region; and (2) that in the red region the relation between warmth and brightness is the reverse of the one implied by Ross's correlations; namely, increasing warmth corresponds with *decreasing* brightness.

With respect to visual judgments of surface colors, Bullough, Collins,

---

[9] F. Stefănescu-Goangă, Experimentelle untersuchungen zur Gefühlsbetonung der Farben, *Psychol. Studien*, 7, 1911, 284-335; R. J. Lewinski, An investigation of individual responses to chromatic illumination, *J. Psychol.*, 6, 1938, 155-160; Ross, *op. cit.*, 141-142, 149, 155, 160, 166.

[10] Ross, *op. cit.*, 183, 155, 166.

Allesch, Tatibana, Tinker, Newhall, and Kimura all found that surface reds were consistently described as warmer than surface blues.[11]

Newhall's study is the most complete. He not only asked for direct temperature ratings and related associations, but had one group of *Ss* list warm and cool objects and another group assign colors to these objects. Thus the temperature association was explicit only for the first group, while the color association was explicit only for the second. His data, from 297 *Ss* judging 50 equispaced Munsell hues, show modes for the 'warmest' at dominant wave-length 610 mμ (red), and for the three 'coolest' at 572 mμ (yellow-green), 552 mμ (green), and 476 mμ (blue). It is not clear, however, how much the familiarity or popularity of these hues determined the precise modes obtained.

With respect to the association of memory-colors with the adjectives 'warm' and 'cool' both Tatibana and Rogers reported that 90% of their *Ss* uniquely associated red and orange with 'warm' and green and blue with 'cool.'[12] Neither of these men kept track of both 'lightness' and 'saturation,' however, hence information concerning the effects of these dimensions on the apparent warmth of surface colors is limited. Allesch surmised from introspective reports that increasing lightness decreases apparent warmth.[13] Newhall commented that some of his *Ss* said dark colors were warmer and light colors cooler; and in Tinker's study, which suggests that surface texture is not a prominant factor in temperature or prefance ratings, white was judged rather less warm than black.[14] It is not possible, however, to know from these studies how much the unrecognized effects of lightness and saturation governed the observed effect of hue.

With respect to saturation, both Tinker and Kimura found that all their colors were ranked warmer than black, gray, or white, which suggests that, in the limit, increasing saturation produces increasing apparent warmth. Collins used two versions of each hue which he spoke of as 'saturated' and as 'weak' and observed that when his *Ss* were asked to assign meanings to 20 two-color charts, 83% of the 'warmth' but only 56% of the 'cool' assignments went to the so called 'saturated' colors. Saturation is, however, frequently confused with lightness and since Collins gives no information about the lightness of his 'saturated' and 'weak' colors the implications of his data are obscure.

---

[11] Edward Bullough, The 'perceptive problem' in the esthetic appreciation of single colors, *Brit. J. Psychol.*, 2, 1908, 406-463; N. Collins, The appropriateness of certain color combinations in advertising, M.A. Thesis, Columbia University, 1924; G. J. von Allesch, Die ästhetische Erscheinungsweise der Farben, *Psychol. Forsch.*, 6, 1925, 1-91; Y. Tatibana, Color feeling of the Japanese: I. The inherent emotional effects of colors, *Tohoku Psychologica Folia*, 5, 1937, 21-46: M. A. Tinker, Effect of stimulus-texture upon apparent warmth and affective value of colors, this JOURNAL, 51, 1938, 532-535; S. M. Newhall, Warmth and coolness of colors, *Psychol. Rec.*, 4, 1941, 198-212; Toshio Kimura, Apparent warmth and heaviness of colors, *Japanese J. Psychol.*, 20, 1950, 36.

[12] Tatibana, *op. cit.*, 27-29; M. E. Rogers, A study of color preferences, M.A. Thesis, Purdue University Library, June, 1950.

[13] Allesch, *op. cit.*, 83-91.

[14] Tinker, *op. cit.*, 533.

The partial regressions of average 'kalt-warm' ratings on hue, lightness, and saturation obtained in the present study are given in Table II. There is a well-defined hue-effect on apparent warmth, independent of lightness or saturation. The regression coefficient and its standard error imply that as hue moves in Munsell steps from red at 5 to blue at 75, average ratings on a seven-point scale from 'warm' to 'kalt' can be expected to shift from at least half a position on the scale to perhaps as much as two positions toward 'kalt'; *i.e.* between 10% and 30% of the way.

The effects of lightness and saturation correspond in direction with the implication of previous work, but are less definitive than the effect of hue. In this study, increasing ratings of 'warmth' correspond with increasing saturation. As in Tinker's study, white is rated colder than black, and confidence-intervals overlap around the lightness-regression coefficient and the white-black estimate. In each case, however, a reasonable confidence-interval (say, two standard errors) around the estimated coefficient of regression includes a possible coefficient of zero.

Weight. Bullough, De Camp, Koch, Monroe, Warden and Flynn, Taylor, Tatibana, Kimura, and Payne all have reported data from which it can be inferred that increasing lightness decreases apparent weight.[15]

Bullough set out to determine whether "the weight-principle itself may, in the case of shade differences, be traceable to either saturation or luminosity";[16] and he concluded from his data that, "luminosity not saturation is the true index of the weight of a color."[17] But in the data he reported there is no evidence at all on the question of saturation, and other research suggests that saturation may in fact have some slight effect. With respect to lightness, Bullough asked his *Ss* which of a pair of two-colored figures they preferred. He used triangles, squares, and circles. The figures differed only in that the lower half of one was the same color as the upper half of the other. The colors differed in lightness, and were both matched and unmatched for hue. About 75% of Bullough's 50 *Ss* preferred figures in which the darker color was in the lower half. This was true whether colors were matched for hue or not. It is from this expression of preference that he infers the effect of lightness on apparent weight.[18]

---

[15] Bullough, On the apparent heaviness of colors, *Brit. J. Psychol.*, 2, 1907, 111-152; DeCamp, *op. cit.*, 356, 363, 366; B. C. Koch, The apparent weight of colors, *Ohio State University Studies: Graduate School Series: Contributions in Psychology*, No. 9, 1928; Monroe, *op. cit.*, 198, 206; Warden and Flynn, The effect of color on apparent size and weight, this JOURNAL, 37, 1926, 398-401; Taylor, *op. cit.*, 241, 243-245; Tatibana, *op. cit.*, 27-29; Kimura, *op. cit.*, 36; M. C. Payne, Apparent weight as a function of color, this JOURNAL, 71, 1958, 725-730.

[16] Bullough, *op. cit.*, 114.

[17] *Ibid.*, 151.

De Camp, Warden and Flynn, Taylor, Kimura, and Payne used a more direct method. They asked their *Ss* to judge how heavy colored boxes 'looked.' Their work is limited to a few colors plus in most instances, black, gray, and white. Only hue and lightness were observed and the effects of these two dimensions were confounded in their analyses. There is agreement, nevertheless, among these five studies which supports the inference of a color-effect. Black was found to 'look' heaviest, followed by red and blue, which in turn were followed by yellow and white, and Payne reported a correlation of 0.94 between apparent weight and reflectance for his six colors. Tatibana asked Japanese students to associate the adjectives 'light' and 'heavy' with 4-in. squares of surface color and obtained essentially the same results.

The only study on color-weight, in which all three of the dimensions of color were controlled, is Monroe's. She ran three sets of comparisons, one for each dimension, and in every case tried to hold the other two dimensions constant. Unfortunately, in the comparison of hues, she was unable to separate the effect of hue from a confounding with lightness. Her experimental method was to ask her *Ss* to move a vertical pointer, as if it were a fulcrum, between two circles of surface color until they "appeared exactly to balance." She then recorded the deviation from the geometrical mid-point between the colored circles. She found that lightness produced an average shift from the mid-point of a 200-mm. line of 7.6 mm. toward the darker color, while saturation produced an average shift of 3.5 mm. toward the less saturated color. From this she concluded that both decreasing lightness and decreasing saturation increased apparent weight. The main criticism of her study, aside from the confounding of hue and brightness, is that her results were based on the judgments of only six *Ss*—three men and three women.

Table II gives the partial regression of average 'leicht-schwer' ratings on hue, lightness, and saturation obtained in the present study. There is a well-defined effect of lightness on apparent weight. The regression-coefficient and its standard error imply that as lightness moves in Munsell units from near black at a value of one to near white at a value of nine, average ratings on a seven-point scale from 'schwer' to 'leicht' can be expected to shift from at least two and a half to perhaps as much as four and a half scale-positions toward 'leicht,' *i.e.* between 40% and 60% of the way. The estimate of the effect of lightness based on white-black difference is very close to the partial regression coefficient based on 45 colors.[19]

---

[18] Kimura observed identical esthetic trends among Japanese students.

The effect of saturation on apparent weight corresponds with Monroe's findings in direction but is rather less in relative magnitude. She reports a saturation-effect about 40% as large as that of lightness while this data suggests a saturation-effect relative to the lightness of color of the order of 10%. Increasing 'leicht' ratings correspond with increasing saturation, but a 95% interval of confidence around the estimated regression coefficient includes a possible coefficient of zero.

SUMMARY

Forty-five colors were rated independently on the semantic differential adjective-pairs 'kalt-warm' and 'leicht-schwer.' The linear regression of average ratings of 'warmth' and 'weight' on hue, lightness, and saturation were reported and previous studies on the apparent warmth and weight of colors were reviewed. The virtues of the present study are (1) a large number of colors were used; (2) observations were focused strictly on visual judgments; and (3) the partial effects of hue, lightness, and saturation were examined explicitly.

With the exception of locating 'cool' modes more precisely and determining whether they are truly modes of temperature or the result of differential color-familiarity, the effect of hue on apparent warmth seems to be established beyond the need for further research. The effects of lightness and saturation, however, while strongly implied, still need additional study. The large residual mean square for 'kalt-warm,' coupled with Ross's data, suggests that it may be worthwhile to investigate the existence of non-linear terms, especially interactions between hue and lightness and hue and saturation, in the relation between color and apparent warmth.

---

[19] A comment about the 'lightness' pun in 'light-heavy' and 'light-dark' is in order. While the presence of this pun in English may be linguistic evidence for the influence of lightness of color on apparent weight, the use of this pun to establish the influence begs the question. The data reported here, however, were collected in German. A common translation of 'light-dark' in German is 'hell-dunkel.' The translation coming closest to this pun is 'licht-dunkel' and 'licht' is distinct in spelling and in sound from 'leicht.' Thus it is difficult to explain the West German data as no more than the manifestation of a 'lightness' pun.

The effect of lightness of color on apparent weight has been found in terms of (1) light-dark design preferences; (2) visual observation of colored boxes and colored squares; (3) make-believe balancings of colored circles; (4) the association of memory-colors; and (5) average adjectival ratings. With the exception of efforts to translate apparent weight into actual units of weight, research on the effect of lightness on apparent weight seems complete.

The effect of saturation, while strongly suggested by the present study and by Monore's data, needs further investigation. Certainly the effect of saturation is substantially less in magnitude than the effect of lightness. As for hue, the present study suggests that its effect is, at most, less than that of saturation.

AMOUNT AND RATE OF DECAY OF VISUAL FIGURAL AFTER-EFFECT AS FUNCTIONS OF TYPE OF INSPECTION-STIMULUS AND INSPECTION-TIME

By Carl P. Duncan, Northwestern University

This experiment has two purposes. First, to determine the effect of inspection-time on both the initial amount of visual figural after-effect developed and on the rate of the decay of the after-effect. Although many studies have seemed to show that amount of after-effect increases (up to a limit) as inspection-time increases,[1] a number of Japanese investigators believe that magnitude of after-effect is the same, regardless of length of inspection-time.[2] According to these workers, increases in inspection-time only slow the rate of decay of the after-effect. After-effects begin to decay the instant the inspection-stimulus is removed, and it usually takes at least a few seconds to measure after-effect. Therefore, according to the Japanese argument, briefer inspection-times yield lesser amounts of after-effect than longer times because the shorter the inspection-time, the greater the dissipation of after-effect before the moment of measurement. In a recent study which did not bear directly on this issue, Carlson reported that he could best account for his data by adopting the Japanese position.[3] In the present experiment, measurements of after-effect are taken at several points in time following removal of the inspection-stimulus. If the Japanese are right, one would expect to find little or no difference in amount of after-effect for different inspection-times when measurement is made as soon as possible after inspection, but as decay-time increases, the decay-curves should gradually diverge.

The second purpose of the study was to determine whether the amount of after-effect varies as a function of certain characteristics of the inspection-stimulus; viz. the width of the stimulus, and whether the stimulus

---

* Received for publication October 14, 1959. Supported by Grant B-792 from the National Institutes of Health, U. S. Public Health Service. Thanks are due Dr. Bruce O. Bergum, Keith Clayton, William Vermillion, and Thomas Ard for their assistance in this study.

[1] Peter McEwen, Figural after-effects, *Brit. J. Psychol., Monogr. Suppl.*, 1958 (No. 31), 1-106.

[2] Moriji Sagara and Tadasu Oyama, Experimental studies on figural after-effects in Japan, *Psychol. Bull.*, 54, 1957, 327-338.

[3] V. R. Carlson, After-effect of a moving pattern, *J. exp. Psychol.*, 58, 1959, 31-39.

is a solid or an outline-figure. Interest in these variables stemmed from Köhler and Wallach's discussion of the distance-paradox (called here the distance-gradient).[4] They presented both data and theory indicating that the peak of the distance-gradient, *i.e.* the distance from the inspection-stimulus at which maximal after-effect occurs, is farther away from a narrow, solid inspection-stimulus, such as line, than from a wide stimulus, such as a solid oblong. The same result occurred when the wide, solid oblong was replaced by the corresponding outline-figure of the same width; the peak of the gradient was farther away from the outline than from the solid figure. Köhler and Wallach pointed out that these findings do not imply that the *amount* of after-effect will be greater with a line or with an outline-figure; although the peak of the distance-gradient may be farther away from the narrow figures than it is from a wide, solid oblong, the latter may be a more intensely satiating stimulus. Thus, the present study also attempts to determine whether after-effect, always measured at a fixed distance from the inspection-stimulus, varies in amount when the inspection-stimulus is a wide or a narrow solid figure, and is a solid or an outline-figure.

METHOD

Stimuli. Four inspection-stimuli were prepared, each drawn in India ink on a separate sheet of white posterboard. All stimuli were 7½ in. tall; this is the vertical distance available in the exposure-field of the Gerbrands tachistoscope used to present the stimuli. The two wide stimuli were ½ in. in width; the narrow two were 5/16 in. in width. One of the wide and one of the narrow stimuli were solid; the other two were closed outlines. The contour of the outline-stimuli was 3/32 in. on all four sides. Thus, the inspection-stimuli permitted study of two values on each of two stimulus-variables: width and type (solid vs. outline).

As *S* viewed a white card bearing an inspection-stimulus in the tachistoscope, he saw a small black fixation-cross in the center of the card with the stimulus to the right of the cross. The distance from the center of the cross to the outside, right edge of the inspection-stimulus was 45/64 in.; this distance was constant for all stimuli.

The test-figure consisted of two black hairlines (actually nylon fish line), 7½ in. long, 1/64 in. wide, on another white posterboard card. The lines were to the left and right of a fixation-cross in the center of the card. The right test-line was fixed at a distance of 47/64 in. from the center of the cross, thus, this line was 1/32 in. to the right of all inspection-stimuli. This distance should permit measurable figural after-effect (apparent displacement of the right line to the right) to occur with all stimuli.

The left test-line was supported by a wooden frame the edges of which fitted into tracks mounted above and below the white card. Card, tracks, and frame were inserted into the preëxposure field of the tachistoscope. Since the preëxposure box was to *S*'s right as he looked into the eyepiece, *S* could easily grasp the end of the wooden frame, which protruded from the box, with his right hand. Sliding the frame into or out of the box made the left test-line move, in *S*'s view, left or right. *S*'s judgment consisted of attempting to set the left test-line the same distance from the fixation-cross as the right line. A pointer mounted on the frame enabled *E* to record *S*'s judgment from a scale fixed to the tachistoscope.

---

[4] Wolfgang Köhler and Hans Wallach, Figural after-effects: An investigation of visual processes, *Proc. Amer. phil. Soc.*, 88, 1944, 269-357. See especially 301-303 and 339-340.

The total pattern formed by the two test-lines when equidistant from the fixation-cross subtended 3°40′ of visual angle. The luminance of the white fields in the tachistoscope, measured at the eyepiece by a MacBeth illuminometer, was 9.7 ml.

Design. Inspection-time was varied by presenting each of the four inspection-stimuli either 15 sec. or 60 sec. Thus, there were eight experimental conditions: two inspection-times combined with two widths and two types of inspection-stimuli. An independent group of *Ss* was tested in each condition. A control group, not presented with the inspection-stimuli, was also tested.

Subjects. The *Ss* were undergraduates assigned to groups in turn. After the experiment was completed, it was found that in three of the experimental groups a few extreme scores were distorting the means and variances of pre-inspection judgments in comparison with the pre-inspection judgments of the other six groups. After discarding these extreme scores, and discarding scores in other groups to produce equal *N*s, there remained 288 *Ss*, 32 in each of the 9 groups.

Procedure. The eyepiece of the tachistoscope was equipped with a white poster-board shutter with three positions: *E* could cover *S*'s right eye only, cover his left eye only, or cover both eyes. When *S* first appeared for the experiment, he was seated at the tachistoscope with the shutter so set that he could not see into the tachistoscope with either eye. The fluorescent ceiling-lights were on, as they were throughout the experiment. (Average luminance in the room was about 7.5 ml., so *S* had to adapt relatively little when looking into the tachistoscope.) *E* pointed out the eyepiece and the wooden frame to *S* and explained that when *S* looked into the eyepiece, he would see two thin lines with a small cross between them. The *S* was told that the right line was stationary and that his job would be to move the wooden frame until he judged the left line was the same distance from the cross as the right line. At this point *E* had *S* look into the eyepiece, uncovered *S*'s right eye to give him a brief look at the test-lines, then closed the shutter again. Then *E* told *S* that he would have to make his adjustments of the left line while looking only at the cross. *S* was told that every time he looked into the tachistoscope throughout the entire experiment, he would see a small cross and it was strongly emphasized that he must never take his eyes from the cross at any time. Finally, *S* was told he would make several adjustments of the left line, that accuracy in adjusting the line was more important than speed, but that it was not worthwhile spending too much time on one adjustment.

S was given eight pre-inspection trials (adjustments). Before each trial, *E* had set the left line clearly nearer or farther away from the cross than the right line (farther on odd-numbered trials, nearer on even-numbered trials). *S* looked into the eyepiece, *E* uncovered his right eye only, and said, "Fixate the cross" at the beginning of every trial. As soon as *S* indicated that he had made his adjustment, *E* immediately closed the shutter and told *S* to let go of the wooden frame. After taking the reading, *E* displaced the left line (starting points for the left line were varied haphazardly), and the next trial began.

After pre-inspection Trial 4, *S* was given a rest (about 20 sec.). If *S* had been taking more than 10 sec. to make an adjustment, *E* suggested that *S* was taking more time than was probably necessary. *S* was then given pre-inspection Trials 5-8. These trials served as the baseline against which to measure figural after-effect.

Following the pre-inspection trials, the *S*s in the experimental groups were given the single inspection-period. *E* told *S* that the next time *S* looked into the eyepiece he would be looking with the left eye and that all he had to do was fixate the cross for a time. *S* was told that after a while the two lines would again be shown and that adjustments would be made as before. The inspection-stimulus for the group to which *S* had been assigned was already in place in the exposure-box. *E* uncovered *S*'s *left* eye, pressed the button to expose the inspection-stimulus, and said, "Fixate the cross." After 15 or 60 sec. of inspection, depending upon the condition to which *S* had been assigned, *E* simultaneously switched the field back to the test-lines, moved the shutter to cover *S*'s left eye and uncover his right eye, and told *S* to fixate the cross and adjust the left line. There then followed 15 additional post-inspection settings of the test-lines, 16 post-inspection trials in all, with a 30-sec. rest after each block of four trials.

The control group was run in the same way as the experimental groups except that for the control *S*s a 60-sec. rest substituted for the inspection-period.

RESULTS

Initial amount of after-effect. For each *S*, the mean of pre-inspection Trials 5-8 was obtained. This value was subtracted from the score on post-inspection Trial 1; the difference is the measure of figural after-effect. A difference with positive sign suggests the presence of after-effect, *i.e.* indicates that *S* set the left line farther to the left to compensate for apparent displacement of the right line to the right.

It is first necessary to determine whether the experimental situation produced significant figural after-effect. For all the experimental *S*s combined (256), the mean after-effect on Trial 1, in units of 1/32 in., was 0.534. The corresponding value for the control *S*s, *i.e.* the difference between the mean of 'pre-inspection' Trials 5-8 and 'post-inspection' Trial 1, was −0.051. The *t*-value for the difference between the over-all experimental mean and the control mean was 2.66, significant at the 1% level. Thus, the inspection-stimuli as a whole produced a significant mean displacement of the right test-line of slightly more than 1/64 in. Since the viewing distance in the tachistoscope was 23 in., this is a displacement of about 2.5′. McEwen estimated that in a similar experimental situation Köhler and Wallach had obtained a maximal displacement of about 3′.[5]

---

[5] McEwen, *op. cit.*, 17; *cf.* Köhler and Wallach, *op. cit.*, 301. Two millimeters in their measurement equal approximately 0.5/32 in. in the present experiment.

The figural after-effect scores for the eight experimental groups were compared by an analysis of variance.[6] The triple-classification analysis yielded terms for inspection-time, width of inspection-stimulus, type of inspection-stimulus, and their interactions. The summary table of the analysis is not presented here because none of the F-values, either for the three main variables or their interactions, was significant. Amount of after-effect on Trial 1 did not differ as a function of any of the three experimental variables.

Since none of the interactions in the analysis of variance was significant, one can obtain two means, one for each of the two values of an experimental variable, by combining groups differentiated on the other two variables. The three pairs of means, each mean based on 128 *S*s, are shown in Table I. By *t*-test the six means were all significantly different from the control mean, −0.051, at the 5% level or better, but as the analysis of variance showed, the means within each pair do not differ.

TABLE I

MEAN FIGURAL AFTER-EFFECT, IN UNITS OF 1/32 IN., ON THE FIRST POST-INSPECTION TRIAL FOR EACH OF THE TWO GROUPS DIFFERENTIATED ON EACH OF THE THREE INDEPENDENT VARIABLES

For each variable, Groups are combined on other variables to yield means based on an N of 128.

Variable	Group	Mean	σ_M
Inspection-time	15 sec.	.559	.112
	60 sec.	.509	.115
Stimulus-width	narrow	.465	.100
	wide	.603	.125
Stimulus-type	solid	.465	.116
	outline	.603	.111

Decay of after-effect. Since it was shown that significant figural after-effect occurred on the first post-inspection trial, decay of after-effect was examined in terms of the raw post-inspection scores, *i.e.* without subtracting the pre-inspection baseline. The post-inspection scores were handled as follows: First, the scores for the 16 trials for each *S* were combined into eight pairs to obtain a mean score for Trials 1-2, 3-4 . . . 15-16; this eliminates fluctuation due to direction-of-adjustment error. Then groups were combined to yield three pairs of decay-curves corresponding to the three pairs of major groups ($N = 128$ for each group) denoted in Table

---

[6] The test for heterogeneity of variance yielded $F_{max} = 2.17$, not significant with 8 and 31 *df*.

I. The graph for one of these pairs of groups, that for 15 sec. vs. 60 sec., is shown in Fig. 1.

Three trend-analyses were done on the decay-curves, one analysis for each of the three pairs of major groups. The scores (trend-points) used in each trend-analysis were those for Trials 1-2, 3-4, 7-8, 11-12, and 15-16. The summary of the analysis for the 15-sec. group vs. the 60-sec. group is given in Table II. The purpose of the trend-analysis is, of course, to determine if there is a significant interaction between the variable (inspection-time in Table II) and trials. Table II shows that the Trials × inspection-time interaction was significant at the 5% level. This supports what one would gather by inspection of Fig. 1; the 15-sec. group shows some decay of after-effect, the 60-sec. group does not. Table II also shows that the over-all means of the groups do not differ (*F* for inspection-time was not significant), and that the term for trials did not quite reach the 5% level of significance. The non-significant term for trials also suggests that one group showed some decay, the other group none. As a further exploration of this point, trend-analyses were done for the 15-sec.

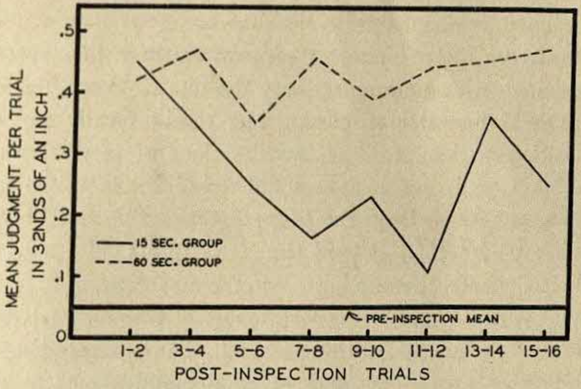


FIG. 1. DECAY OF AFTER-EFFECT AS A FUNCTION OF INSPECTION-TIME

TABLE II

SUMMARY OF ANALYSIS OF VARIANCE FOR TREND OF THE 15-SEC. AND 60-SEC. GROUPS OVER POST-INSPECTION TRIALS

Source	*df*	Mean square	*F*	*p*
Inspection-time	1	10.74		
Between *S*s in same group	254	14.40		
Trials	4	1.13	2.13	
Trials × Inspection-time	4	1.32	2.49	.05
Pooled *S*s × Trials	1016	.53		

and 60-sec. groups separately. For the 15-sec. group, the *F* for trials was 4.20 ($p < 0.01$, 4 and 508 *df*); for the 60-sec. group, the *F* for trials was less than one.

Summaries of the trend-analyses for the Wide vs. Narrow groups, and the Solid vs. Outline groups are not presented. None of the terms in either analysis was significant. Therefore, the graphs showing the decay-curves for these groups are also not presented.

DISCUSSION

The first purpose of the present study was to check the Japanese claim that differing inspection-times affect only the rate of decay, not the initial amount, of visual figural after-effect.[7] The results clearly supported the Japanese position, at least for the two durations of inspection-time used here. The 15-sec. and 60-sec. groups did not differ in initial magnitude of after-effect, as measured on the first post-inspection trial. Over the 16 post-inspection trials (total elapsed time: 5-7 min.), after-effect decayed significantly in the 15-sec. group but not in the 60-sec. group.

The finding that rate of decay was inversely proportional to duration of inspection is not surprising; this possibility has been suggested before,[8] but the finding of no difference in initial magnitude of after-effect as a function of inspection-time was actually more clear-cut than had been expected. The Japanese criticism of experiments showing after-effect increasing as a function of inspection-time was that measurements were not taken the instant the inspection-stimulus was removed, so the measurements were confounded by differential rates of decay. It is surprising that this effect did not occur in the present experiment. The *S*s were allowed to take their own time, within limits, to make a judgment. Most *S*s took from 5-10 sec. per judgment. Thus, it might have been expected that even on the first post-inspection trial, the 15-sec. group would show less after-effect than the 60-sec. group, but this did not occur.

The present findings on inspection-time do not, of course, indicate that *all* inspection-times will yield the same amount of after-effect immediately after inspection. It is possible that both 15 and 60 sec. are on the asymptote of a curve which does in fact show initial magnitude of after-effect to increase as inspection-time increases. A systematic investigation of inspection-durations within the range of a fraction of a sec. to 5-10 sec. is needed.

Even if it turns out to be generally true that initial amount of contour-displacement is constant for all inspection-times, whereas rate of decay of displacement is not, such facts would not necessarily be difficult to account for by neural theory of satiation. One could assume that with a given kind of stimuli at a given luminance, any amount of satiation will produce a constant (the maximum for that situation) amount of displacement. There seem to be no serious objections to such an assumption. At the same time, continued exposure to the inspection-stimulus at a fixed location in the visual field could very well increase the amount or intensity of satiation of the neural tissue. More intense satiation, produced by longer inspection-durations, would presumably take longer to dissipate, so slopes of decay-curves would be inversely related to intensity of satiation.

---

[7] Sagara and Oyama, *op. cit.*

[8] C. P. Duncan, On the similarity between reactive inhibition and neural satiation, this JOURNAL, 69, 1956, 227-235.

The second purpose of the present experiment was to determine whether different sizes and kinds of inspection-stimuli produce different initial amounts of after-effect. The stimulus-dimensions used, wide vs. narrow, solid vs. outline, were selected on quasi-theoretical grounds, viz. there is some evidence indicating that these dimensions significantly affect the peak of the distance gradient.[9] The results were clear-cut. There were no differences in initial magnitude of after-effect, nor for that matter in rate of decay, attributable to the stimulus-variations. Perhaps it is also worth noting that there was no interaction, in initial amount of after-effect produced, between the two stimulus-dimensions, although this finding does not seem intuitively impressive because of the failure to find significant main effects.

In concluding that the stimulus-variations had no differential effect, one precaution is suggested: the conclusion should be limited quite strictly to the specific parametric conditions of the experiment. The difference in width between the wide (½ in.) and narrow (5/16 in.) oblongs was not great. The distance of the right, fixed test-line from the inspection-stimuli (1/32 in., about 5′ of visual angle) was fixed. If it is true that the distance gradient varies with the stimulus-dimensions used here, the decision to use the 1/32-in. separation, and in particular the decision to keep the separation constant, may have been inadvisable.

Finally, a methodological point may be worth mentioning. It is the writer's impression that the tachistoscope (Gerbrands and similar types) is not a particularly good device for studying visual figural after-effects. In a previous study, no after-effect was obtained in a tachistoscope quite similar to the one employed in the present study.[10] Over-all magnitude of after-effect in the present experiment was not large. The writer believes, without much evidence, that a short viewing distance (23 in. in the tachistoscope used here), a small visual field (here, $7\frac{1}{2}$ in. square) surrounded by a deep black background, or perhaps the combination of both makes it more difficult to obtain measurable after-effects.

---

[9] Köhler and Wallach, *op. cit.*

[10] C. P. Duncan, Visual figural after-effect as a function of stimulation interval and decay time, *Percept. mot. Skills*, 8, 1958, 203-206.

SUMMARY

The experiment dealt with the effects of several stimulus-variations on visual figural after-effect. Two hypotheses were tested. The first was that initial amount of after-effect would not vary with duration of the inspection-period but that rate of decay of after-effect would be inversely proportional to inspection-time. The second was that initial amount of after-effect would vary as a function of width and type of inspection-stimulus. There were two durations of the inspection-period (15 sec. and 60 sec.), and two widths and two types (solid or outline) of inspection stimuli. These conditions were arranged in a factorial design with 32 *S*s assigned to each of the eight conditions. The experimental *S*s attempted to adjust two hairlines equidistant from a central fixation-point both before and after fixating an inspection-stimulus. An additional control group of 32 *S*s went through the same procedure without viewing the inspection-stimuli. Amount of contour-displacement, or figural after-effect, was measured by the difference between pre-inspection and post-inspection adjustments.

In comparison with the control group of *S*s, the experimental group showed significant initial figural after-effects, the mean contour-displacement being about 2.5′ of visual angle. There were no differences in initial amount of after-effect produced by any of the experimental conditions or their interactions. Rate of decay of after-effect over 16 post-inspection trials was inversely related to inspection-time, but did not vary as a function either of width or type of inspection-stimulus. Taken together, the results supported the first hypothesis but not the second.

TRANSFER OF LEARNING WITH DIFFERENCES IN ASSOCIATIVE VALUE AND IN MANIFEST ANXIETY

By LUCIANO L'ABATE, Washington University School of Medicine

The empirical grounds for the role of manifest anxiety as a motivational variable in complex verbal learning and transfer are still tenuous and controversial.[1] According to theoretical assumptions underlying manifest anxiety, where transfer from previous learning facilitates a response, the performance of anxious *S*s should be superior to that of nonanxious *S*s. Where an old response interferes with the acquisition of a new response, a reversal should take place: the performace of nonanxious *S*s should be superior to that of anxious *S*s. Evidence to support this prediction was found only for men under three conditions of transfer, while women behaved in opposite fashion. The absence of a reasonable explanation for the sex-differences found then in learning and transfer made necessary a more rigorous replication, attempted in the present study.[2] Another difficulty involved has been the absence of control over the associative values and the lack of a wide range of strength for the response-tendencies tested. The associative value of the response-tendencies is one of the most crucial factors in verbal learning and transfer.[3] If it is assumed that performance in verbal learning is a multiplicative function of habit-strength

---

* Received for publication May 26, 1959. Supported by Grant M-1373 from the National Institute of Mental Health to East Carolina College, Greenville, North Carolina. This report was written during the author's tenure as USPHS Post-Doctoral Fellow at Michael Reese Hospital, Chicago. The author is indebted to Dr. Helen Heath for assistance.

[1] J. A. Taylor, The effects of anxiety level and psychological stress on verbal learning, *J. abn. soc. Psychol.*, 57, 1958, 55-60; W. M. Nicholson, The influence of anxiety upon learning: Interference or drive increment? *J. Pers.*, 26, 1958, 303-319; D. H. Kausler and E. P. Trapp, Methodological considerations in the construct validation of drive-oriented scales, *Psychol. Bull.*, 56, 1959, 152-157.

[2] Luciano L'Abate, Transfer and manifest anxiety in paired-associate learning, *Psychol. Rep.*, 2, 1956, 119-126; Manifest anxiety and the learning of syllables with different associative values, this JOURNAL, 72, 1959, 107-110.

[3] R. H. Day, Relative task difficulty and transfer of training in skilled performance, *Psychol. Bull.*, 53, 1956, 160-168; J. M. Vanderplas, Transfer of training and its relation to perceptual learning and recognition, *Psychol. Rev.*, 65, 1958, 375-385; George Mandler and E. H. Campbell, Effect of variation of associative frequency of stimulus and response members on paired-associate learning, *J. exp. Psychol.*, 54, 1957, 269-273; B. J. Underwood and R. W. Schulz, *Meaningfulness and Verbal Learning*, 1960, Chap. 2, 3, and 4.

and drive, then various combinations of associative values should interact with manifest anxiety.[4]

By controlling as much as possible for sex-differences and for associative value, the hypotheses underlying the present investigation were as follows: four nonsense-syllables are designated as A, B, C, and D. Pairs of these syllables are used as stimulus- and response-members, (S_A–R_B) respectively, in paired-associate learning. (1) If the stimulus-syllable is changed while the response-syllable remains the same (S_C–R_B), anxious *S*s should show greater facilitation than nonanxious *S*s in connecting the old response to the new stimulus; (2) if the response is changed while the stimulus remains the same (S_A–R_D), anxious *S*s should show greater interference in learning the new response than nonanxious *S*s; (3) if both the stimulus and the response are changed (S_C–R_D), anxious *S*s should learn the new task less readily than nonxious *S*s; and (4) the associative value for each pair, as predicted from the individual associative values of the syllable themselves, should interact multiplicatively with manifest anxiety to produce greater facilitation or interference for syllables with higher associative value, depending on the transfer-condition.

METHOD

Subjects and design. Four hundred and eighty undergraduate students were selected as *S*s on the basis of their scores on Taylor's Scale of Manifest Anxiety. These *S*s scored in the upper and lower quartiles of the distribution to divide anxious (HA) from nonanxious (LA) *S*s. They were drawn from a more homogeneous student population than that used in a previous study to exercise a greater degree of control over possible sex-differences.[5] To fit the experimental design shown in Table I and to balance sex- and anxiety-level, 12 subgroups of 40 *S*s each were formed for each of the 12 learning- and transfer-conditions. Each subgroup consisted of 10 HA and 10 LA men and 10 HA and 10 LA women. The experimental design was a factorial one, employing two conditions for sex, two for anxiety-level, three transfer-conditions, and four lists.

Materials and procedure. Nonsense-syllables rated 100 and 0 in associative value were selected from Glaze's tables. They were divided into four lists, according to conventional criteria of selection.[6] Two lists, each of 10 syllables, had associative values of 100. Two lists had associative values of 0. These lists were combined into paired-associates and served in different groups as either stimuli or responses, according to the experimental design shown in Table I. It will be noted that each of the transfer-conditions involved the substitution of a new set of syllables for the stimulus-member of the first pair, for the response-member, or for both, and the substitution was always of a list of different associative values than the original list. They were presented in three different orders to avoid serial effects.

---

[4] K. W. Spence, *Behavior Theory and Conditioning*, 1956, 226; K. W. Spence, I. E. Farber, and H. H. McFann, The relation of anxiety (drive) level to performance in competitional and non-competitional paired-associates learning, *J. exp. Psychol.*, 52, 1956, 296-305; K. W. Spence, John Taylor, and Rhoda Ketchel, Anxiety (drive) level and degree of competition in paired-associates learning, *ibid.*, 52, 1956, 306-310.

[5] L'Abate, *op. cit.*, 1956, 125.

[6] E. R. Hilgard, Methods and procedures in the study of learning. In S. S. Stevens (ed.) *Handbook of Experimental Psychology*, 1951, 540; L'Abate, Recognition of paired trigrams as a function of associative value and associative strength, *Science*, 131, 1960, 984-985.

*S*s were assigned to each experimental subgroup without *E*'s knowing about their anxiety-levels. They were not aware of the purpose of the experiment. Learning and transfer-lists were administered individually on a Lafayette memory drum set to expose the stimulus-term alone for 2.4 sec. and the stimulus-response pair for a like time. The *S*s memorized as many paired-associates as possible for a fixed number of 10 learning- and 10 transfer-trials, using an anticipation-method.

TABLE I

EXPERIMENTAL DESIGN: NUMBERS INDICATE GLAZE'S PERCENTAGE OF ASSOCIATIVE VALUE OF WORDS IN EACH LIST

Learning First task	Transfer conditions: Second task stimulus-change	response-change	stimulus-response change
S_A–R_B	S_C–R_B	S_B–R_D	S_C–R_D
100–100	0–100	100–0	0–0
0–100	100–100	0–0	100–0
100–0	0–0	100–100	0–100
0–0	100–0	0–100	100–100

TABLE II

MEAN NUMBER OF CORRECT RESPONSES AND *SD*s IN LEARNING AS A FUNCTION OF ASSOCIATIVE VALUES, SEX, AND ANXIETY-LEVEL

($N=30$ per cell)

Lists S_A—R_B	Men HA Mean	Men HA *SD*	Men LA Mean	Men LA *SD*	Women HA Mean	Women HA *SD*	Women LA Mean	Women LA *SD*	Totals Mean	Totals *SD*
100–100	23.07	13.69	30.00	14.32	33.30	13.67	33.20	14.43	30.02	14.03
0–100	14.57	7.63	15.00	8.83	14.53	8.75	14.10	9.31	14.55	8.63
100–0	17.60	8.40	18.40	10.66	17.60	14.50	18.53	9.89	18.03	10.86
0–	6.20	3.54	5.77	3.29	7.23	7.27	7.17	6.16	6.59	5.06
Total	15.36	8.34	17.29	9.27	18.16	11.05	18.37	9.95		

RESULTS

Learning. The mean number of correct responses for the first 10 learning-trials is summarized in Table II. On the average, women learned slightly faster (18.26) than men (16.32), but this difference was mostly due to the much lower mean (23.07) of the HA men on List 100–100. This discrepancy was actually produced by 10 HA men in the response-interference condition (S_A–R_D). Besides the small differences between men and women, more marked differences were produced by the differences among lists, the List 100–100 producing the greater number of correct responses and list 0–0 the lowest, as expected. List 100–0 produced a much greater mean number of (unexpected) correct responses than List 0–100.

The correct responses during 10 learning-trials were subjected to analyses of variance. Sex yielded a significant main effect ($F = 4.16$, $<0.025\ p < 0.05$) and a significant interaction with lists ($F = 3.22$, $0.05 < p < 0.10$). Lists yielded an extremely high main effect ($p < 0.001$). Because of the significant sex-interaction, which was due to the greater number of correct responses in the women, learning-scores were analyzed separately for men and women. There were extremely high main effects of lists in both sexes ($p < 0.001$). Transfer-conditions yielded a main effect ($F = 3.50$, $0.025 < p < 0.05$) for men only, because of the lower number of correct responses for HA men in the response-interference condition already mentioned. No other significant main effect or interactions were found.

Transfer. The complex experimental design precluded the use of more conventional measures of transfer. Two adjustments were used, therefore, to reach a measure of transfer that would be more appropriate to the experimental design and to the differences in mean number of correct responses found during the 10 learning trials. The first adjustment was made to control for the unequal difficulty of the various lists. The amount of transfer was obtained by taking the difference between transfer and original learning on the same lists by comparable groups. The measure was the deviation y', in terms of the individual number of correct responses, from the mean number of correct responses in original learning for the same kind of *S*s on the same list. As an example, 30 HA men gave an average of 23 correct responses while learning the 100–100 list. This number was subtracted from the score achieved in transfer by each HA man on List 100–100.

The second adjustment was made to eliminate as much individual variability as possible and attempted to control for the amount of learning of each *S*. It was assumed that the greater the amount of learning on the first list, the greater the transfer on the second list. The individual score y predicted on the basis of the regression of transfer on the original learning, as measured by the deviation y' and the number of correct responses in learning for all *S*s (HA men and women, LA men and women) on the same list, was subtracted from deviation y'. For example, the individual deviation-score from the mean of the original learning for List 100–100 was correlated to the number of correct responses for List 0–100 achieved by the same individual in learnings (Table I), thus giving the regression for each group. The difference ($y'-y$) between the obtained y' score and the predicted score resultant from the second adjustment (y) gave the individual transfer-score. Such a transformation reduced all lists and all transfer-conditions to zero value, thus eliminating main effects for both variables, but allowing us to test their interactions with anxiety-level and sex. The results of this transformation are summarized in Table III.

The mean transfer-scores for Condition S_C–R_B increase for HA *S*s as associative value increases, while LA *S*s show a decline in the opposite direction. The difference between HA and LA *S*s on List 100–100 is statistically significant ($t = 2.55$, $0.01 < p < 0.05$). As associative value decreases, LA *S*s show a trend toward improvement. Nonetheless, the total mean transfer-scores for the over-all transfer-condition indicate the superiority of HA *S*s, regardless of sex-differences. In the response-interference conditions S_A–R_D, the absolute values of the mean transfer-scores for all *S*s decrease to a minimal difference, except for the slight increase on List 100–100. Women conformed to the prediction of greater over-all interference, with HA women showing a decrease in transfer-scores. Men did not conform to the prediction, and HA men were still superior to LA men, although a tendency to reverse this trend with degree of low associative value (0–0) is apparent. The expected reversal in the performance of HA and LA *S*s occurred under conditions of stimulus-response interference S_C–R_D. The LA *S*s showed a greater degree of facilitation than HA *S*s, especially on List 0–100 in both sexes.

TABLE III

MEAN TRANSFER SCORES (y'–y) AND *SD*s AS A FUNCTION OF ASSOCIATIVE VALUES, TRANSFER CONDITIONS, SEX, AND ANXIETY LEVEL
($N = 10$ per cell)

Transfer conditions	Associative values	Men				Women			
		HA		LA		HA		LA	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
S_C–R_B	100–100	8.75	15.22	−3.29	10.99	.37	6.69	−5.81	9.86
	0–100	1.00	14.38	.52	7.55	3.41	11.00	−4.97	9.33
	100–0	−3.06	10.55	−1.96	7.47	5.30	8.54	− .27	14.19
	0–0	− .30	1.98	2.60	4.17	−1.26	3.64	−1.05	7.89
	Total (N=40)	1.60	10.53	− .53	7.54	1.95	6.56	−3.02	10.32
S_A–R_D	100–100	9.98	9.99	−4.50	14.05	−6.40	10.58	.92	8.57
	0–100	−2.81	4.56	−1.82	7.43	3.81	11.85	.79	11.50
	100–0	−3.63	3.31	−2.26	4.84	2.63	9.61	3.30	7.25
	0–0	1.90	4.08	.41	2.14	−1.65	4.39	− .64	4.17
	Total (N=40)	1.36	5.48	−2.04	7.11	− .40	9.11	1.09	7.87
S_C–R_D	100–100	1.95	16.20	−1.70	16.86	−2.29	17.39	2.06	14.39
	0–100	−5.56	6.04	3.30	6.27	− .63	12.89	2.89	13.91
	100–0	1.26	12.15	2.03	8.96	−1.82	13.37	−1.49	7.19
	0–0	.54	1.81	.37	5.53	− .21	4.11	− .70	3.25
	Total (N=40)	− .45	9.05	1.00	9.40	−1.24	11.94	.69	9.68

Analysis of variance for the mean transfer-scores yielded a significant third-order interaction between anxiety, sex, and lists ($F = 10.07$, p

< 0.001) and significant second-order interactions between sex and lists ($F = 2.74$, $0.025 < p < 0.05$), and between anxiety and transfer conditions ($F = 2.81$, $0.05 < p < 0.10$). The third-order interaction between sex, anxiety-level, and lists was also due to the performance of women, who conformed to the prediction of greater response-interference for the HA women, while men behaved contrary to the prediction. Separate analyses of variance for men and women were done because of the significant third-order interaction. Two second-order interactions were found: between anxiety and lists for men ($F = 5.84$, $p < 0.001$) and between anxiety and transfer-conditions for women ($F = 2.92$, $0.05 < p < 0.10$).

DISCUSSION

In terms of associative values, the learning results of the present investigation confirm previous findings concerning the positive relationship between ease of learning and increasing associative values. List 100–0, however, shows an unexpected, over-all superiority over List 0–100. The same differences between lists of varying associative values were replicated in transfer, again with the exception of Lists 0–100 and 100–0. Such an almost paradoxical superiority may only be initial and may occur only within the first 10 trials. It may be due to perceptual factors as yet unclarified and to different shapes in the acquisition-curves.[7]

A partial but encouraging verification of the hypotheses underlying the present investigation was obtained in terms of the transfer-measure employed. Men verified the hypothesis that anxiety-level would interact significantly with associative values. The verification of the hypotheses based on the predicted relationship between anxiety-level and transfer-conditions, although supported by a significant interaction in the total sample, depended primarily on the performance of women, especially in the response-interference condition. Sex-differences, therefore, cannot be ignored. Precautions were taken to insure a selection of *Ss* from a more homogeneous background than the sample of the original study. The women were highly similar to the men in terms of intelligence, motivational level, and socio-economic and geographic background. This persistent difference over three studies raises a question as to the different motivational needs of men and women that has not as yet been considered in relation to manifest anxiety. Different motivational variables in men and women may be selected by the Taylor Scale of Manifest Anxiety. For instance, women may show a greater degree of neuroticism and less defensiveness then men in answering the questionary, as has been found in children.[8]

The first hypothesis, that greater response-facilitation should result whenever a high level of manifest anxiety is present, was verified in Condition S_C–R_B. Furthermore, such a transfer-condition verified the fourth hypothesis—that various degrees of associative values should interact with anxiety-level according to a multiplicative relationship. The degree of transfer for HA *Ss* increased noticeably as the degree of associative value increased. The opposite trend occurred with LA *Ss*. Where associative values were very low, on List 0–0, the differences between HA and LA *Ss* disappeared completely. In fact, under conditions of low associative values, one would expect LA *Ss* to perform better than HA *Ss*.[9]

---

[7] Vanderplas, *op. cit.*, 375-385.

[8] Leona Tyler, *The Psychology of Human Differences*, 1956, Ch. 10; L'Abate, Personality correlates of manifest anxiety in children, *J. consult. Psychol.*, 24, 1960, 342-348; S. B. Sarason *et al.*, *Anxiety in Elementary School Children: A Report of Research*, 1956, Ch. 9.

The lack of verification for the second hypothesis in the response-interference condition S_A–R_D for men should not come as a surprise and should not be taken as being contradictory to theoretical expectations. Aside from considerations about sex-differences, such a condition has been known to contain a far lesser degree of response-interference than the one assumed. That a certain degree of response-interference does take place is shown by the decrease in the absolute values of the mean transfer-scores for all *Ss*. Complete response-interference is a difficult condition to obtain experimentally. This difficulty is mainly due to other variables not manipulated in the present experiment, such as: a greater number of trials in learning and the degree of inter-list similarity. Even a mean number of 19 trials in learning was not sufficient to produce a satisfactory amount of interference for all *Ss* in the same transfer-condition in the original study.[10] Then, women behaved contrary to the prediction, while in the present study women performed in the predicted direction. If it is assumed that a greater number of trials in learning is necessary to produce a satisfactory amount of response-interference, then anxiety may be still acting as an energizer to increase the degree of transfer, as in the case of HA men on List 100–100.[11]

The third transfer-condition, S_O–R_D, confirms the third hypothesis and adds support to the fourth hypothesis. The expected reversal in the degree and direction of transfer indicates that the extent of stimulus-response interference, as measured by the transformation employed, may be even greater than in the previous transfer-condition. In fact, *Ss* must acquire completely new stimulus- and response-connections. The performance of HA *Ss* is poorer than that of LA *Ss* because of the interference of the stimulus- and response-connections already acquired and energized during learning. The lack of a greater difference between HA and LA groups on List 100–100 is due to the insufficient number of interfering tendencies in learning, since very few correct responses could have been acquired by either group in learning List 0–0. List 0–100 still illustrates the possibility that a multiplicative relationship between associative values and anxiety-level may exist whether facilitative or interfering response-tendencies are present.

If all three transfer-conditions are considered to be on a continuum of progressive response-facilitation, response-interference, and even greater stimulus-response interference, then the performance of HA *Ss* becomes increasingly poorer than the performance of LA *Ss*. Similar findings have been reported in regard to the effects of anxiety-level on concept-formation and the initial phases of problem-solving.[12]

---

[9] C. D. Spielberger, L. D. Goodstein, and W. G. Dahlstrom, Complex incidental learning as a function of anxiety and task difficulty, *J. exp. Psychol.*, 56, 1958, 58-61.

[10] L'Abate, *op. cit.*, 1956, 119-126.

[11] C. C. Spiker and R. B. Holton, Associative transfer in motor paired-associate learning as a function of amount of first-task practice, *J. exp. Psychol.*, 56, 1958, 123-132.

[12] C. W. Romanow, Anxiety level and ego involvement as factors in concept formation, *J. exp. Psychol.*, 56, 1958, 166-176; S. J. Blatt and I Stein, Efficiency in problem-solving, *J. Jsychol.*, 48, 1959, 193-213.

In conclusion, when anxiety is present, it energizes all response-tendencies. When response-interference is also present, anxiety will energize interfering tendencies as well. Where stimulus-facilitation is present, anxiety interacts with high associative value to improve performance of HA *S*s and with low associative values to improve the performance of LA *S*s. Where the interference is due to both stimuli and responses, regardless of associative values, LA *S*s perform better than HA *S*s.

SUMMARY

The hypotheses retested in the present investigation were that a high level of manifest anxiety should produce greater response-facilitation, greater response-interference, and greater stimulus-response interference than a low level of manifest anxiety. An additional hypothesis concerned the multiplicative relationship that should exist between associative values and manifest anxiety.

Four hundred and eighty undergraduates were selected from the upper and lower quartiles on the distribution of scores achieved on Taylor's Scale of Manifest Anxiety. To fit the experimental design of the study, the *S*s were divided into 12 subgroups, each consisting of 40 *S*s counterbalanced for sex and anxiety-level (10 anxious men and 10 anxious women, 10 nonanxious men and 10 nonanxious women). All *S*s were instructed individually to learn lists of paired-associate nonsense-syllables differing in degree of associative value. (100–100; 0–100; 100–0; 0–0) for a fixed number of 10 trials. Then, they learned another list of paired-associates for a fixed number of 10 trials in one of the three transfer-conditions.

By using a complex measure of transfer, derived from the difference between the obtained and the predicted number of correct responses in transfer, most of the hypotheses were supported. Although the primary hypotheses were supported by a significant interaction between anxiety-level and transfer-conditions for the total sample, women appeared to be mainly responsible for such an interaction. The additional hypothesis concerning the multiplicative relationship between anxiety level and associative strength was verified by a significant interaction between anxiety-level and lists in men only. The persistent sex-difference made it impossible to verify the hypotheses underlying the present investigation to a greater degree of significance and consistency.

A FURTHER EXAMINATION OF THE SCHAFER-MURPHY EFFECT

By JOHN F. SANTOS, The Menninger Foundation, and E. A. GARVIN
Griffiss Air Force Base, Rome, New York

Since Schafer and Murphy reported that monetary rewards and punishment affected the perception of reversible profiles,[1] their study has been repeated several times but with ambiguous results. Rock and Fleck, for example, in a modified replication, failed to obtain autistic effects, whereas Jackson, following Schafer and Murphy's procedure obtained corroborative results.[2] He failed to do so, however, when he employed Rock and Fleck's procedure. The difference between the two procedures involved the method of presenting the stimulus-objects. Other studies that have since been made also confirm and negate Schafer and Murphy's findings.[3]

Various criticisms have been made of Schafer and Murphy's study.[4] The most serious are: (1) their sample of *Ss* is too small to permit generalization; (2) they confounded profile and direction by presenting a given face always in the same direction (facing right or left) for all *Ss*; (3) they arbitrarily eliminated the results of the last 16 test-trials from their statistical analyses. The first and second criticisms have been considered in the studies of Jackson and of Solley and Long, and the last by Rock and Fleck and by Jackson.[5] The present experiment, therefore, extends the previous work by employing isolated as well as interacting rewards and punishments and by using an additional control group. This study also takes account of the criticisms listed above.

Method: Subjects. Twenty-six boys and 22 girls between 10 and 13 yr. of age were used in the experiment. Seventeen additional *Ss* were discarded because they

---

* Received for publication July 24, 1960.

[1] Roy Schafer and Gardner Murphy, The role of autism in a visual figure-ground relationship, *J. exp. Psychol.*, 32, 1943, 335-343.

[2] Irvin Rock and F. S. Fleck, A re-examination of the effect of monetary reward and punishment on figure-ground perception, *J. exp. Psychol.*, 40, 1950, 766-776; D. N. Jackson, A further examination of the role of autism in a visual figure-ground relationship, *J. Psychol.*, 38, 1954, 339-357.

[3] D. E. P. Smith and Julian Hochberg, The effect of "punishment" (electric shock) on figure-ground perception, *J. Psychol.*, 38, 1954, 83-87; C. M. Solley and John Long, Perceptual learning versus response set learning, *Percept. mot. Skills*, 8, 1958, 235-240.

[4] Schafer and Murphy, *op. cit.*, 338-343.

[5] Jackson, *op. cit.*, 340-342, 350; Solley and Long, *op. cit.*, 236-237; Rock and Fleck, *op. cit.*, 769-772; Jackson, *op. cit.*, 344-348, 350-352.

failed to associate correctly given letters with given profiles. All *Ss* were students in the public or parochial school systems in New Orleans, Louisiana. They were chosen from the upper half of their classes on the basis of grades and normal eyesight.

Apparatus. The apparatus was similar to that used by Rock and Fleck.[6] It consisted of an SVE Speed-i-o-scope (tachistoscope) set at 'bright' illumination and 1/3-sec. exposure, which was attached to the barrel of an SVE Miniature Projector (Model AK, 300 W.). Three 35-mm. slides were used to project the stimuli which were single faces or their ambiguous combinations (see Fig. 1). The ambiguous stimuli contained both faces within a circle 24 in. in diameter and the single stimuli merely omitted one or another half of the full circle to achieve a relatively unambiguous face. The dimensions of the screen were 48 in. × 48 in.

Procedure. The procedure was similar to that used by Rock and Fleck.[7] There were a training phase, a test phase, a post-experimental test of learning, and an interview. Only one pair of Schafer and Murphy's single faces (referred to here as *A* and *C*, respectively),[8] the mirror images of these (referred to here as *B* and *D*, respectively), and their combination into ambiguous faces (referred to here as *A/C* and *B/D*, respectively) were used. The *S* was seated 10 ft. from the projection screen and the projector was situated 13 ft. from the screen. During the training, *Ss* learned to identify two of the single faces (*A* and *C* or *B* and *D*) by a letter. One *E* said aloud the letter identifying the face to be presented, then flashed the face and required *S* to repeat the letter. Under some of the conditions, 10¢ was given every *S* at the beginning of training and, during training, 1¢ was given (reward) or taken away from him (punishment) in conjunction with the presentation of certain faces. Each *S* was shown both faces of a pair (*A* and *C* or *B* and *D*) 10 times each during each training period.

Table I indicates the experimental groups and their treatment with respect to the conditions of reward and punishment. The Reward–Punishment Group (*R–P*) was always rewarded for one face and punished for the other, and thus ended with the same amount of money with which it had begun (10¢). The Reward–No Reward Group (*R–NR*) always received reward for one profile, but nothing for the other, and finished with 20¢. The Punishment–No Punishment Group (*P–NP*) invariably lost money for one face but not for the other and therefore was left with nothing. For the Reward–Punishment Control Group (*R–PC*), rewards and punishments were given in a random manner to both profiles so that these *Ss* (like those in the *R–P* Group) ended with 10¢ each.

Twelve *Ss* were randomly assigned to each of the four groups; hence age and sex were roughly equivalent from group to group. As shown in Table I, there were four subgroups of three *Ss* each, so that profile (*A* and *C* or *B* and *D*), direction (*A* and *B* or *C* and *D*), and conditions of reinforcement (*R, NR, P, NP*) were counterbalanced. The *R–PC* Group was used to determine whether the mere presence of rewards and punishments during training (apart from their systematic relationship to certain stimuli) might result in effects which in turn would influence the frequency of occurrence of certain perceptual responses to the ambiguous stimuli.

---

[6] Rock and Fleck, *op. cit.*, 768.
[7] Rock and Fleck, *op. cit.*, 767-768.
[8] Schafer and Murphy, *op. cit.*, 336-337.

Following training, the *S*s were told that they would be shown the same profiles as before, but this time they would be in an enclosed rather than half a circle and that now they would not be rewarded or punished. The *S* was instructed to identify verbally the profile that was flashed with an appropriate letter. A forced-choice situation was imposed. *E* alerted *S* by saying "ready" before each presentation. Sixteen exposures of ambiguous profiles (*A/C* and *B/D*) were given to all *S*s. On eight of these, the faces appeared in the same position and direction as during training and on the other eight exposures the faces were mirror images, *i.e.* 'reversed' in position and direction from that used during training.

A test of learning followed in which *S*s were told that the faces would be presented again as they were originally, and they were to identify them with letters as before. Every *S* was given three presentations of each of the profiles used in training and, if he failed to give six correct identifications, he was eliminated as an *S* on the assumption that either he had never learned the appropriate identifications or had forgotten them.

TABLE I

NUMBER OF RESPONSES TO TEST-FIGURES FOR GROUPS AND SUBGROUPS

	Groups							
Comparison	Reward & No-Reward (*R–NR*)		Punishment & No-Punishment (*P–NP*)		Reward & Punishment (*R–P*)		Control (*R–PC*)	
	R	*NR*	*P*	*NP*	*R*	*P*	*RP*	*RP*
A vs. *C*	28	20	0	48	23	25	26	26
C vs. *A*	16	32	33	15	24	24	24	25
B vs. *D*	24	24	23	25	20	28	17	31
D vs. *B*	30	18	22	26	25	23	24	24

After the test of learning, *S*s were questioned regarding the certainty of their responses, their feelings toward reward and punishment, their preferences for the two faces, and whether they ever perceived two faces during the presentations of the ambiguous profiles.

Results. The frequencies of the alternative responses of individual *S*s were converted into exact cumulative probabilities based on the binomial distribution, and these probabilities in turn were converted into standard normal deviates. The *z*-scores of individual *S*s within each group were then summed algebraically and divided by the square root of the number of *S*s, and this quotient was reconverted into a probability-value. With minor modifications, this procedure follows Stouffer's method of combining observed *p*-values.[9]

The analyses of responses for individual *S*s within the four groups indicated that Conditions *R–NR*, *R–P*, and *R–PC* failed to influence perceptual responses to a significant degree ($\bar{z} = 0.32$, 0.57, and 0.78 respectively). On the other hand, Condition *P–NP* produced differences (in favor of *NP* reports) which, for the total group, were significant beyond the 2% level of confidence ($\bar{z} = 2.40$).

---

[9] Frederic Mosteller and R. R. Bush, Quantitative techniques, in G. Lindzey (ed.) *Handbook of Social Psychology,* 1, 1954, 329.

There were significant differences between the two sets of responses, *A,B* and *C,D* in Group *P–NP* ($\bar{z} = 4.47$, $P < 0.001$, $A,B > C,D$) but not in Groups *R–NR*, *R–P*, and *R–PC* ($\bar{z} = 0.80$, 0.29, and 0.50 respectively). There is, therefore, no consistent preference for either face.

It has been suggested that the influence of reinforcement on perception should be most evident on the first test-trial because this measure is uncontaminated by the effects of previous responses or extinction.[10] The frequency of first responses to the ambiguous stimuli was, therefore, checked with respect to the experimental conditions employed within a given group. In Group *R–NR*, seven *S*s gave first responses indicating perception of faces that had been rewarded, while the five remaining *S*s identified faces that had not been rewarded. In Group *P–NP* the first responses of only three *S*s identified punished faces, while those of nine *S*s identified faces that had not been punished. The first responses of Group *R–P* showed that five of these *S*s perceived rewarded faces and seven *S*s perceived punished faces. Obviously, the first responses present a picture quite similar to the one established when all of the responses made during the test-phase are considered.

In the interview, the *S*s were questioned regarding their certainty in making their responses. It was found that only 3 *S*s in Group *P–NP* guessed, whereas 8 *S*s in Group *R–NR*, 10 in Group *R–P*, and 6 in Group *R–PC* did so. Another interesting fact is that 10 *S*s in Group *P–NP* expressed concern with monetary loss, while only 6 of those in the other groups admitted to any such feelings. Finally, Group *P–NP* was the only one in which there was a definite relationship between expressed preference for a profile and the one which was most often reported. In every case the non-punished face was preferred.

The number of *S*s in the four groups who did not learn the designations of the faces ranged from two in Group *P–NP* to five in Group *R–P*.

The training procedure was intended to increase the strength or probability of occurrence of certain perceptual responses. Since reinforcements were variable with respect to position but consistent with respect to profiles (except in Group *R–PC*) it was anticipated that facial preferences should be strengthened more than positional preferences and this in turn would play an important role in determining what was perceived when ambiguous stimuli were presented. Using a criterion of less than five exceptions, it was found that facial preferences decreased from four in Group *P–NP* to one in Group *R–PC,* whereas positional preferences ranged from seven in Group *R–PC* to four in Group *R–NR.* Group *P–NP* was relatively low with five instances of positional preferences. These results are contrary to expectations in showing more positional than facial preferences, but the relatively high frequency of the former and the relatively low frequency of the latter among the control *S*s as compared to experimental *S*s are in keeping with expectations.

---

[10] Robert Sommer, Perception and monetary reinforcements: II. The effects of rewards and punishments in the visual modality, *J. Psychol.*, 42, 1956, 143-148.

Discussion. The results of this study are in agreement with some of the previous findings but not with others.[11] With the bewildering array of corroborative and conflicting evidence that exists regarding this phenomenon it is difficult to draw any hard and fast conclusions, but a few facts are obvious. It is apparent that the Schafer-Murphy effect is not as general as might have originally been expected. Even a reasonably good replication may not prove successful, since the phenomenon seems to depend to some extent upon factors in the experimental situation which are not fully controlled or specified, such as the relationship between *E* and *S* and the motivational impact of the reinforcements.[12] It is also possible that more potent reinforcements than those employed in the present experiment are required to bring about changes in expectancies, searching and scanning behavior, and a degree of attention great enough to exert an important influence on what is perceived.

While there is little evidence in the present study for autism in perception, a few facts point in this direction. There are the striking results of the first three *S*s in Group *P–NP* who reported the neutral profile to the complete exclusion of the *P* face, but unfortunately there is no indication of a consistent trend in this direction among the remaining *S*s. As a matter of fact, there are some disturbing reversals. On the other hand, the data obtained in the interviews along with a consideration of facial and positional preferences and the ability to associate a given name with a given face, suggest that autistic effects certainly came closer to being realized in Group *P–NP* than in any of the others.

---

[11] Jackson, *op. cit.*, 345-347, 351; Rock and Fleck, *op. cit.*, 768-773.

[12] C. M. Solley and Mary Engel, Perceptual autism in children: The effects of reward, punishment and neutral conditions upon perceptual learning, *J. gen. Psychol.*, 1960, 97, 77-92.

SUMMARY

This study reports a further examination of the Schafer-Murphy effect. A larger sample of *S*s was used than in previous experiments, along with additional experimental and control conditions. The over-all results suggest the absence of autistic effects. Only the condition in which one figure was associated with punishment and the other was not yielded significant results, and these were primarily due to the responses of one subgroup. Further analyses of the experimental data and interviews with the *S*s indicate that autistic effects were most nearly approximated in this group. A consideration of evidence from a number of studies necessitates the conclusion that the Schafer-Murphy effect may be obtained only under highly specific conditions, and even then uncontrolled motivational and interpersonal variables may have an important influence on the experimental results.

EXTINCTION OF CONDITIONED VERBAL RESPONSES

By ALAN S. DEWOLFE, VA Hospital, Hines, Illinois

Considerable interest has been aroused by the demonstration that verbal behavior can be acquired in essentially the same manner as are instrumental responses of animals, that is, by 'reinforcing' the designated verbal behavior with some stimulus.[1] Studies have also shown that level of performance during acquisition is positively related to the relative frequency of reinforcement, and that responses acquired under partial reinforcement fall off less rapidly during extinction, findings which are paralleled in other learning situations.[2] In addition, reinforcing a conflicting response has been demonstrated to produce more rapid extinction than simply omitting the reinforcement.[3]

In the present study, the influence of various schedules of reinforcement and methods of extinction has been studied. Three schedules of reinforcement were used during acquisition (100%, 75%, 50%). The procedures used to decrease the frequency of response during extinction were: negative reinforcement of the acquired response (*NR*); simply omitting reinforcement of any kind, or standard extinction (*SE*); and reinforcing a conflicting response (*RCR*). Since schedule of reinforcement and method of extinction were varied simultaneously, the relationship between these two variables could be studied. On the basis of past studies, particularly the study by Cohen, Kalish, Thurston, and Cohen, and on the assumption that verbal conditioning operates according to the same principles as more traditional instrumental conditioning, it was expected that extinction would be more rapid with the latter two methods, *i.e. NR* and *RCR,* than in the standard procedure (*SE*) in which reinforcement is simply omitted. With

---

* Received for publication August 26, 1959. This study, based on an M.S. thesis submitted to the Graduate School of Northwestern University, was directed by Dr. Janet A. Taylor.

[1] Leonard Krasner, Studies of conditioned verbal behavior, *Psychol. Bull.,* 55, 1958, 148-170.

[2] F. H. Kanfer, The effect of partial reinforcement on acquisition and extinction of a class of verbal responses, *J. exp. Psychol.,* 48, 1954, 424-432; M. J. Peterson, Verbal response strength as a function of cultural frequency, schedule of reinforcement, and number of trials, *ibid.,* 52, 1956, 371-376; Mark Spivak and John Papajohn, The effect of schedule of reinforcement on operant conditioning of a verbal response in an autokinetic situation, *J. abn. soc. Psychol.,* 54, 1957, 213-217.

[3] B. D. Cohen, H. I. Kalish, J. R. Thurston, and Edwin Cohen, Experimental manipulation of verbal behavior, *J. exp. Psychol.,* 47, 1954, 106-110.

respect to schedule of reinforcement, it was anticipated that while performance might be positively related to percentage of reinforcement during acquisition, greater resistance to extinction would be found in the groups that had been partially reinforced.

METHOD

Subjects. Ss were 180 men and women from undergraduate classes in psychology at Northwestern University. Fifteen Ss were assigned to each of 12 groups in order of their appearance to be tested.

Materials and apparatus. A set of 160 stimulus-cards, 3 × 5 in., was prepared, with a different verb in the past tense typed in upper-case letters at the center of each. Only verbs listed in the most frequently used 1000 words of the Thorndike-Lorge word-frequency tabulation were used.[4] Each of the 160 cards contained, in addition, the 6 personal pronouns (I, we, you, he, she, they) in upper-case letters on a single line below the verb. The order of pronouns on the card varied systematically. To counteract the possibility that any particular verbs made conditioning of one class of pronouns easier, the cards were divided into two packs of 80 cards each that were made equal in the order of the pronouns. One pack was used for acquisition-trials and the other for extinction-trials. The order of presentation of the packs was reversed for alternate *Ss* within each of the groups. Within each pack, the order of presentation was randomized by shuffling the cards before their presentation to each *S*. An exposure-stand, 9 in. high, with a 2 in. × 4 in. opening through which the *Ss* viewed the cards, hid *E*'s actions while he tabulated responses.

Procedure. *S*'s task was to form a sentence beginning with one of the six pronouns and containing the verb on the card. The *Ss* were divided into 12 groups in a 3 × 4 factorial design using the three schedules of reinforcement as the row variable and the four methods of extinction as the column variable. Thus, during the acquisition phase (Trials 1–80) a set of four groups, one to receive each method of extinction in the extinction-phase, was reinforced on a given schedule. The 100%-groups were reinforced every time *S* used 'I' or 'we,' the 75%-groups were reinforced on the first three of each four such responses, and the 50%-groups were reinforced on the first and then alternate responses. The word "good" spoken by *E* in a flat, unemotional tone was the reinforcement in all cases during this phase.

During the second phase of the experiment (Trials 81–160), three groups, one having received each schedule of reinforcement, continued to receive reinforcement at the same rate as in the acquisition-phase. These groups constituted the controls. Of the remaining nine groups, one from each schedule of reinforcement was assigned to each of the three procedures for extinction. The groups with standard-extinction (*SE*) received no reinforcement of any kind during the extinction-phase. In the groups with conflicting responses (*RCR*), men were reinforced by "good" when they said "he" and "they," while women of these groups were reinforced when they said "she" and "they." The groups with negative reinforcement (*NR*) were negatively reinforced after responding with "I" or "we" by "huh uhh" spoken by *E* in a gutteral grunt. The reinforcement for the latter two groups was given after every appropriate response during the extinction-phase.

---

[4] E. L. Thorndike and Irving Lorge, *The Teacher's Wordbook of 30,000 Words*, 1944.

Each *S* was tested individually in a single session of approximately 30 min. After being seated at the table, *S* was told the study was an investigation of how people make up sentences and was instructed to make up a sentence using the given verb and one of the pronouns as the first word. Every *S* was given two practice-cards containing a verb and four common first names. The 160 experimental cards were then presented for a maximum of 5 sec. each. If at the end of that period no response had been made, *E* attempted to prod *S* with the statement, "Just say the first thing that comes to mind." Only the standard intertrial interval elapsed between the last acquisition- and the first extinction-trial and *S* was given no indication of a change from one phase to another. Upon completion of the extinction-phase, *S* was asked how he decided which word at the bottom to use, which ones he chose most frequently and why, whether anything *E* said influenced his choice of pronouns, and what he thought the experiment was all about.

RESULTS AND DISCUSSION

Acquisition. The 80 acquisition-trials were scored in 5 blocks of 16 trials each. For each set of 60 *S*s having the same percentage of reinforcement, the number of critical responses increased across the five blocks. The difference between the mean score on the first block and the last block was highly significant. There was on the other hand, no significant difference as a result of differences in treatment, neither on total scores nor on the final block. Thus, there was no evidence to suggest that percentage of reinforcement influenced acquisition of the critical response, contrary to expectation. Further, there were no significant differences among the groups corresponding to the extinction-procedures to be used. This suggested that the groups to be compared in extinction were well matched for performance in acquisition.

Extinction. Performance on the second (extinction) phase was also scored in five 16-trial blocks. The control groups that received continuing reinforcement (*CR*) showed the expected continued increase in rate of responding and remained above the three extinction-groups. This may be seen from the curve labelled *CR* in the left half of Fig. 1. The groups receiving differing treatments during extinction diverged and remained separated after the first block of trials. The standard-extinction (*SE*) groups showed the least extinction and the negative reinforcement (*NR*) groups the greatest drop in response-rate.

An analysis of the influence of the schedule of reinforcement during acquisition reveals that the three schedules, although they did not produce differences in performance during acquisition, clearly did produce differences during the extinction-phase. The scores for the groups subjected to the three extinction-procedures were combined. The groups that had 50% reinforcement remained at almost the same level throughout the entire extinction-phase. The groups that received 100% reinforcement dropped off most sharply, while the groups that had 75% reinforcement remained intermediate on the last four blocks of extinction-trials. These results are shown in the right half of Fig. 1.

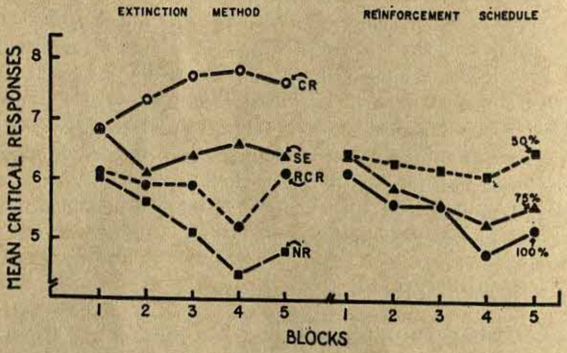


FIG. 1. MEAN CRITICAL RESPONSES MADE PER 16-TRIAL BLOCK IN THE EXTINCTION-PHASE WITH METHOD OF EXTINCTION AS PARAMETER IN LEFT HALF AND SCHEDULE OF REINFORCEMENT IN RIGHT HALF

The data on schedules of reinforcement of the control groups were omitted.

TABLE I

ANALYSES OF VARIANCE FOR CRITICAL RESPONSES DURING EXTINCTION

	All Groups			*CR* Groups omited		
Source	*df*	*MS*	*F*	*df*	*MS*	*F*
Trials (*T*)	4	4.75	1.22	4	14.50	3.63*
T×Extinc. Meth. (*EM*)	12	10.76	2.74*	8	7.25	1.81†
T×Rein. Sched. (*RS*)	8	1.67	0.42	8	2.88	0.72
T×*EM*×*RS*	24	2.79	0.72	16	1.50	0.38
Pooled *Ss*×*T*	672	3.89		504	4.00	
EM	3	201.00	4.04*	2	87.50	2.20
RS	2	9.50	0.19	2	41.00	1.03
EM×*RS*	6	38.20	0.77	4	10.50	0.26
Between *S*s	168	49.70		126	39.80	

* $p<0.01$
† $0.05<p<0.10$

To test the significance of these findings, two analyses of variance were performed. The number of critical responses in each block of trials was taken as the datum, and a 4 × 3 × 5 analysis for repeated measures was done, which is summarized in Table I. The only significant main effect was that of the methods of extinction (*EM*). The trials-by-methods interaction was however, also significant.

Since the rate of response of the control groups continued to increase while it was decreasing for the groups undergoing extinction, the presence of these groups inflated the *F* for methods of extinction and also obscured the effects of trials. Therefore, a second analysis of variance for repeated measures was performed with the data from the *CR* groups omitted. In this analysis, the *F* for trials was now significant. The *F*s for method of extinction and for the trials-by-method interaction, which were significant in the earlier analysis, now, however, no longer quite reached significance at the 5% level.

The assumption that continued extinction-trials would bring the methods of extinction to statistical significance as a variable seems tenable. The results of an unpublished doctoral dissertation by Klein, found in a survey of the literature in verbal conditioning published after completion of the present study, strengthened this assumption.[5] Klein used the same experimental method and the same types of extinction, so that the present study is an extremely close replication of Klein's study. Klein also used 80 extinction-trials and found the same consistent but not statistically significant differences among the three methods. In the second part of his study, he continued the extinction for 180 trials and found the differences among methods were highly significant. Since both the present study and Klein's results suggest that later extinction-trials would show greater significance, the data from the last two extinction-blocks were selected as the most representative of the expected effect of extinction-trials and an analysis of variance was performed on these data. In this analysis, method of extinction was a significant variable ($p < 0.05$). The reinforcement-schedule used during acquisition again did not yield significant differences, but the F of 1.90 had a $p < 0.20$, which was the nearest that this variable had approached significance.

The failure to find that the schedule of reinforcement acted as a significant variable in both acquistion and extinction, particularly in the former, is contradictory to the results of previous studies in this area with respect to this variable.[6] Three factors might help account for the lack of significant effects of the schedule of reinforcement on the rate of extinction. The rate of responding was low; the groups were equal in both rate of responding and number of critical responses in the acquisition phase; and the two additional methods of extinction (*NR* and *RCR*) were more efficient than the standard method, which is characteristically the only one used. These three factors may have been sufficient to mask the effect of the schedules. The differences observed are in the expected direction and increase over trials. The results are not diametrically opposed to the significant results of other studies, but the variable was certainly weak in this situation.

---

[5] Sandor Klein, Conditioning and extinction of operant verbal behavior in neuropsychiatric hospital patients. Unpublished Doctoral dissertation, University of Indiana, 1954.

[6] See references in Footnote 2, above.

To test the effect of awareness on performance, the data from the interviews were used as the basis for classing the *Ss* into three groups differing in the degree of awareness each *S* expressed. The 'aware' group ($N = 38$) contained those who correctly and definitely stated the relationship between the reinforcement and at least one of the critical responses; the 'unaware' group ($N = 83$) were those who gave no hypothesis, or an incorrect one; while the 'partially aware' group ($N = 14$) were those who were indefinite or partially correct. A statistical comparison of the aware and unaware groups was made on the basis of their scores for acquisition and extinction separately. The groups did not differ significantly in either phase as determined by *t*-tests.

SUMMARY

The study was designed to investigate the effect of two variables, schedule of reinforcement during acquisition and method of extinction, on the rate of extinction of a conditioned verbal response. *Ss* were 180 men and women. They were reinforced after 100%, 75%, or 50% of the instances in which they chose a particular pronoun to form a sentence. Their rate of responding with the critical pronouns increased significantly during the acquisition-phase, but there was no significant difference among the groups with varying proportions of reinforcement. Each of the three groups with a given schedule of reinforcement was subdivided into three subgroups that were subjected to different methods of extinction. The groups that were given non-reinforcement or standard extinction showed the slowest decline, reinforcement of a conflicting response gave a greater decrement, and negative reinforcement the largest. A control group from each level of reinforcement received continued reinforcement during the second phase, and continued to improve their performance.

A GROUP-TEST FOR ASSESSING HAND- AND EYE-DOMINANCE

By HERBERT F. CROVITZ, VA Hospital, Durham, North Carolina, and KARL ZENER, Duke University

Scattered evidences that laterality-characteristics of the subject play a role in perceiving have existed for some time. An early series of studies by Dallenbach and co-workers showed that differences in handedness were related to specific patterns of attributive clearness in visual space.[1] In a set of recent independent studies which followed, both hand- and eye-dominance have been related to performance in a variety of visual perceptual tasks in which position or direction in space are of experimental concern. These are indicated in the following review.

Takala reported a series of perceptual experiments in which differences in hand- and eye-dominance were associated with differences in the amount and direction of lateral anisotropies.[2] Gaffron discussed the relation of quadrantic and lateral anisotropies in the phenomenal analysis of pictures to hand- and eye-dominance.[3] Crovitz, Daston, and Zener introduced a task of localization in which a tendency to localize an odd element on the right or the left of a square stimulus-group appeared to be related to handedness or eyedness, or both,[4] while Crovitz showed that in this task, when there is no acuity-difference between the eyes, *S*s who sight with the right eye tend to localize the odd element to the left of the stimulus-group and *S*s who sight with the left eye tend to localize it to the right of the group.[5]

Other investigators interested in the differential effects of right- and left-handedness have found consistency among right-handed *S*s but marked heterogeneity among left-handed *S*s. A significant relationship between directionality of *GSR* reactivity and images of body-variables has been found in right-handed *S*s but not in left-handed *S*s. This has been discussed by Fisher and Abercrombie.[6] Wishner and Ship-

---

* Received for publication October 19, 1960.

[1] K. M. Dallenbach, Position *vs.* intensity as a determinant of clearness, this JOURNAL, 34, 1923, 282-286; R. S. White and K. M. Dallenbach, Position *vs.* intensity as a determinant of the attention of left-handed observers, this JOURNAL, 35, 1924, 267-269; A. M. White and K. M. Dallenbach, Position *vs.* intensity as a determinant of the attention of left-handed observers, this JOURNAL, 44, 1932, 175-179.

[2] Martti Takala, Asymmetries of the visual space, *Ann. Acad. Sci. Finl.*, 27, 1951, Ser. B. (2), 1-150.

[3] Mercedes Gaffron, Right and left in pictures, *Art. Quart.*, 13, 1950, 312-331.

[4] H. F. Crovitz, Paul G. Daston and Karl E. Zener, Laterality and a phenomenon of localization, *Percept. mot. Skills*, 9, 1959, 282.

[5] H. F. Crovitz, Patterns of relative localization of an odd element within a visual grouping as a function of laterality characteristics and tendencies to eye movement. Unpublished Doctoral dissertation, Duke University, 1960.

[6] Seymour Fisher, Body image and asymmetry of body reactivity, *J. abn. soc.*

ley investigated the differential directionality of autokinetic movement in right- and left-handed *S*s, but while they found a tendency for movement to the right among right-handed *S*s, a left-handed group was not consistent in the direction of autokinesis.[7]

It is clear that accurate tests of handedness are required for classification of *S*s. Otherwise *S*s of mixed types will be classified as right- and left-handed to the confusion of studies made upon them. Further, since laterality-effects may be relatively weak determinants of perceptual differences, there is a need for a method of identifying, quickly and efficiently, the *S*s with extreme lateralities. Research either into differences in laterality among groups, or on problems in which functional relationships between laterality and an aspect of performance occur only in certain groups, has been handicapped by the difficulty of selecting the *S*s, particularly when extreme *S*s are sought. Furthermore, additional screening may be needed to determine that the *S*s selected are comparable on other relevant aspects. Effects of laterality may operate at high levels of perceptual integration which may also be affected by other factors. For example, phoria or differential visual acuity may be important, thus requiring further screening of laterality-groups. This is brought to attention by Crovitz's study, in which the direction of lateral localization was found to be related to the direction of an acuity difference favoring one eye.[8] When such additional tests are desirable, the labor of arriving at differential laterality-groups of adequate size may be prohibitive. This is particularly true when the selection of *S*s is conducted on an individual basis. Prior initial screening in a group-situation may, however, be used to locate *S*s in the less frequently occurring groups, as well as *S*s with marked lateralities. Without such screening for *S*s with marked lateralities, ambiguous or negative results may be unnecessarily obtained.[9]

This report describes simple group methods of screening *S*s with respect to hand- and eye-dominance which we have found useful as a first step in selecting *S*s for experimentation on problems concerning the relation of lateralities to perception.

Over a few years, 1569 students in general psychology at Duke University were given screening tests for handedness and eyedness. The test for handedness follows.[10]

TEST FOR HANDEDNESS

Answer the following questions carefuly. Imagine yourself performing the activity described before answering each question. Answer by drawing a circle around the

---

Psychol., 57, 1958, 292-298; Seymour Fisher and Joseph Abercrombie, The relationship of body image distortions to body reactivity gradients, *J. Pers.*, 26, 1958, 320-329.

[7] Julius Wishner and Thomas Shipley, Jr., Direction of autokinetic movement as a test of the "Sensory-Tonic-Field" theory of perception, *J. Pers.*, 23, 1954, 99-106.

[8] Crovitz, *op. cit.*, 19.

[9] In the Wishner and Shipley study, it is possible that this might have occurred, since the criterion of left-handedness was similar to Items 1, 2, 3, 8 and 9 in the test described below, with only half the 'left-handed' *S*s giving results on all of these in a left-handed direction.

[10] Our thanks are due Donald Freedheim, William Groman, and George Greenberg for assisting in the collection of the data and also to Elaine K. Crovitz for assisting in scoring and tabulating the results.

appropriate set of letters appearing to the left of each question whose meanings is:[11]

Ra = right hand always.	Lm = left hand most of the time.
Rm = right hand most of the time.	La = left hand always.
E = both hands equally often.	X = do not know which hand.

(1) Ra Rm E Lm La X: is used to write with.
(2) Ra Rm E Lm La X: to hold nail when hammering.
(3) Ra Rm E Lm La X: to throw a ball.
(4) Ra Rm E Lm La X: to hold bottle when removing top.
(5) Ra Rm E Lm La X: is used to draw with.
(6) Ra Rm E Lm La X: to hold potato when peeling.
(7) Ra Rm E Lm La X: to hold pitcher when pouring out of it.
(8) Ra Rm E Lm La X: to hold scissors when cutting.
(9) Ra Rm E Lm La X: to hold knife when cutting food.
(10) Ra Rm E Lm La X: to hold needle when threading.
(11) Ra Rm E Lm La X: to hold drinking glass when drinking.
(12) Ra Rm E Lm La X: to hold tooth brush when brushing teeth.
(13) Ra Rm E Lm La X: to hold dish when wiping.
(14) Ra Rm E Lm La X: holds tennis racket when playing.

Every item is scored on a 5-point scale. On Items 1, 3, 5, 7, 8, 9, 11, 12 and 14, Ra is scored '1'; Rm, '2'; E, '3'; Lm, '4'; and La, '5.' All other items (2, 4, 6, 10, 13) are scored in the reverse fashion. Items marked X are prorated. The highest possible right-handed score is 14, and the highest left-handed score is 70.

TEST FOR EYEDNESS

A group-test for sighting dominance, described by Greenberg, who demonstrated its usefulness in a study in which a relationship was found between head-tilt and eye-dominance, was used in this study. Greenberg describes his procedure as follows:[12]

> *S* was instructed to sit erect and fixate a point placed upon the blackboard before him. While fixating this point, he was instructed to bring a pencil, which until then had been held vertically with his right [left] hand at his nose into line with the fixation-point. He was then told to close his right [left] eye and to note whether the pencil was still in line with the fixation-point or had shifted to the right or to the left of it.

Eight trials are given every *S* in the group test; two each with a specified eye covered and with the pencil in a specified hand. When the right eye is closed, a report of the pencil remaining in line is classed as a left-eyedness report, while a report of the pencil jumping to the right is classed as a right-eyedness report. When the left eye is closed, a report of the pencil remaining in line is classed as a right-eyedness report while a report of the pencil jumping to the left is classed as a left-eyedness report. If fewer than six of the eight trials are classed as right [left] eyedness, *S*'s eyedness is classed as ambiguous.

RESULTS AND DISCUSSION

These tests have been administered to as many as 250 *S*s in a room at a time. The success of the testing depends, in part, on the skill of the tester in keeping control of the group. The *S*s must be cautioned to keep their heads erect; to move only with specific instructions; and to refrain from practicing between trials. Verifying individual tests of eyedness were given to 127 *S*s. Of these, 22 were found to have been misclassified in the group test. Some *S*s classed as ambiguous with respect to eyedness in the group were found to be right- or left-eyed in individual testing and others who had been classed as right- or left-eyed in the group were found to be ambiguous with respect to eyedness in individual testing. No *S* classed as right- [or left-] eyed in the group was found to be left- [or right-] eyed in individual testing. Nonetheless, the group test of eyedness should be viewed as a screening device rather than as a substitute for further verifying tests of eyedness.

---

[11] For a similar test, see June E. Downey, Types of dextrality and their implications, this JOURNAL, 38, 1927, 321 f.

[12] George Greenberg, Eye-dominance and head-tilt, this JOURNAL, 73, 1960, 149-151.

An assessment of the degree of eyedness may be made in the group test, although such assessment is necessarily approximate. Of the *S*s classed as right- or left-eyed in the group, most used the same eye in sighting on eight of the eight trials. Of the *S*s classed as right-eyed, 89% used the right eye on eight trials, 5% on seven trials, and 6% on six trials. Of the *S*s classed as left-eyed, 84% used the left eye on eight trials, 10% on seven trials, and 6% on six trials. Of all *S*s in the group tests, 18.3% were classed as ambiguous with respect to eyedness. It seems likely that this percentage is not a true estimate of the prevalence of ambi-eyedness, since it would be inflated by the performance of *S*s who were confused or did not follow instructions scrupulously.

The results from the group test of handedness indicate the usefulness of this test in identifying extreme *S*s. Figs. 1 and 2 present the frequency distributions of scores on this test for men and women respectively.

In each case the distribution is sharply skewed. For women the mode is the most extreme right-handed score (14), while for men it is less extreme (18). These figures may prove useful in giving normative data to which performance of new *S*s may be compared.

The scores of *S*s reporting themselves as right-handed, when this question was added to the test for 1059 *S*s, ranged from 14 to 44, and the scores for *S*s reporting themselves as left-handed ranged from 23 to 70. The concentration of scores along the handedness-dimension for *S*s reporting themselves as right- and left-handed is given in Table I.

Assuming that the scores on this test reflect location on an underlying handedness-dimension, these ranges strongly suggest that a more differentiated estimate of handedness, such as that reported here, would permit a more discriminative selection of *S*s in research related to laterality than

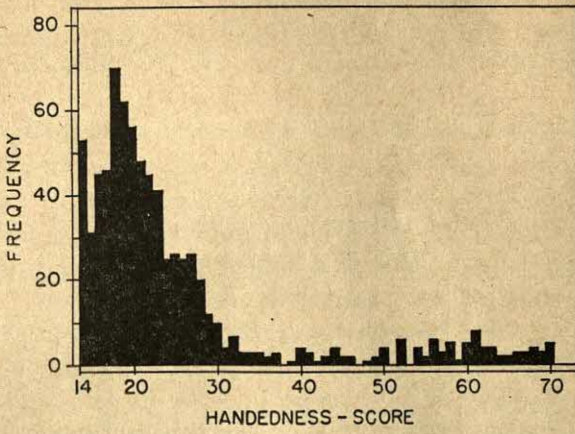


FIG. 1. DISTRIBUTION OF THE SCORES MADE BY THE MEN
($N = 768$)

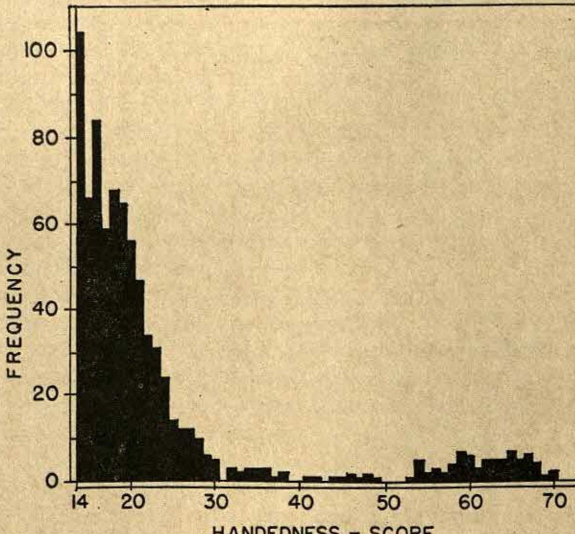


FIG. 2. DISTRIBUTION OF THE SCORES MADE BY THE WOME
($N = 801$)

merely *S*'s report of his handedness or of the hand which is used in writing, especially since all *S*s who classified themselves as right-handed also answered item (1) ("which hand is used in writing") by circling Ra (or Rm) and all *S*s who classified themselves as left-handed circled La (or Lm).

The greatest potential usefulness of these screening tests may be in facilitating selection of groups of *S*s for experiments whose major focus is the specification of the role of the laterality-factor in perception and the clarification of their modes of action. Greenberg's study, however, demonstrates the usefulness of group-testing not only for screening purposes, but as a direct tool for research.[13]

TABLE I

PERCENTAGE OF *S*S WITH HANDEDNESS-SCORES FALLING IN SELECTED RANGES

Handedness by self-report

Scores	Right	Left	Scores	Right	Left
14–20	67	0	41–50	0	15
21–30	31	2	51–50	0	35
31–40	2	8	61–70	0	40

---

[13] Greenberg, *op. cit.*, 150.

THE COMPLEXITY OF CONCEPT-ATTAINMENT

By LISE WALLACH, Wellesley College

According to common sense, it is clear that there is something like 'complexity,' and that when something gets 'complicated' it gets harder. But what is complexity? When is one thing more complicated than another? The purpose of the present study was to attempt to understand complexity in the area of concept-attainment. To illustrate the problem, suppose that instead of our usual system of traffic signals, one had to learn that red lights meant 'stop' on north-south streets, and 'go' on east-west streets. This seems more complicated, and one would expect it to be harder to learn. What is there about it that makes it seem more complicated and harder?

Two factors that have been discussed in the literature on concept-attainment clearly seem relevant.[1] They are (1) the structure of the concept, and (2) the number of attributes involved. If red meant 'stop' on north-south streets, and 'go' on east-west streets, then the meaning of red would be contingent upon the direction of the streets. Both the addition of direction as a relevant attribute and the contingency of the relations here seem to contribute to complexity.

May, then, the complexity of attaining a concept be defined in terms of the number of relevant attributes and their structural relations? Although this seems to be in the right direction, an analysis in terms of attributes may not be adequate.

Suppose, for example, that in a typical experiment on concept-attainment, *S*s learn to call figures which have a half-circle, a straight line, and a sharp angle 'vec,' and others 'not vec.' Patterns like those in the top row of Fig. 1 they would call 'vec,' and in the bottom row, 'not vec.'

One *S* might proceed by noting the presence or absence of half-circles, straight lines, and sharp angles. Another might simply note a similarity in all the 'vecs' without ever concerning himself separately with half-circles, lines, and angles. One intuitively believes that the first *S*'s concept involves more than that of the second. This is, however, not simply more *attributes,* at least in any sense in which attributes are a function only of the stimulus-situation, and do not vary with the *S*. In 'objective' terms, exactly the same attributes might be involved for both *S*s—as would be indicated by their calling the same stimuli 'vec' and 'not vec,' no matter how many new stimuli they were given. Such considerations suggest that the variables

---

* Received for publication November 20, 1959. Based on a dissertation submitted to the University of Kansas in partial fulfillment of the requirements for the Ph.D. degree. My thanks are due Professor Alfred L. Baldwin for his advice and criticism, and to Drs. Edward L. Wike, Rhea Diamond, and Michael A. Wallach for their helpful suggestions.

[1] See J. S. Bruner, J. J. Goodnow, and G. A. Austin, *A Study of Thinking,* 1956.

which matter in complexity are not objective, in the sense of remaining invariant with changes in the *S*.

Recent work which has been summarized by Miller has indicated that a particular kind of subject-dependent units is critical in the process of memorization.[2] The amount of material an *S* can memorize seems to be an almost direct function of the number of these 'cognitive units' or 'chunks.' It seems very possible that these units are also what is relevant in concept-attainment. The complexity of concept-attainment, then, may depend on the number of cognitive units and the structural relations among them.

Cognitive units are still not too well understood. If one views cognition as a kind of coding, a cognitive unit may be regarded as a unit of a code that a person may utilize. In what cognitive units a stimulus-pattern will be coded depends in part upon the person, and in part upon the stimulus, including the gestalt factors of proximity, similarity, and smooth continuation. The unit character involved in cognitive units is the same as the oneness of the 'complex ideas' of the British associationists.

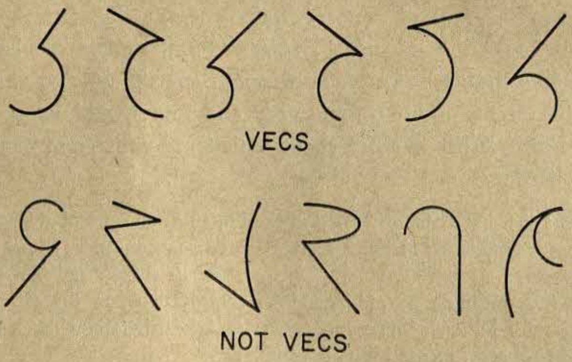


FIG. 1. 'VEC' AND 'NOT VEC' PATTERNS

This brings us back to the problem of concept-attainment. Concepts themselves are cognitive units, and, after the earliest period of life, most concepts seem to be acquired by the unitizing of groups of other cognitive units. This is the process with which this paper is concerned—and what is being suggested is that the difficulty of this unitizing depends on the number of cognitive units to be united and the structure in which they are to unite.[3] The following experiment was designed to test some implications of this statement.

---

[2] G. A. Miller, The magical number seven, plus or minus two: some limits on our capacity for processing information, *Psychol. Rev.*, 63, 1956, 81-97; Information and memory, *Sci. Amer.*, 195, 1956, 42-46.

[3] The acquisition of the basic cognitive units, which will not be discussed here, is, of course, also an extremely important problem.

DESIGN AND PROCEDURE

Sixty students enrolled in the summer session at Cornell University served as the *Ss* in this study in which the concepts differed in structure and number of cognitive units.

The materials were two sets of eight cards containing a variety of patterns. The

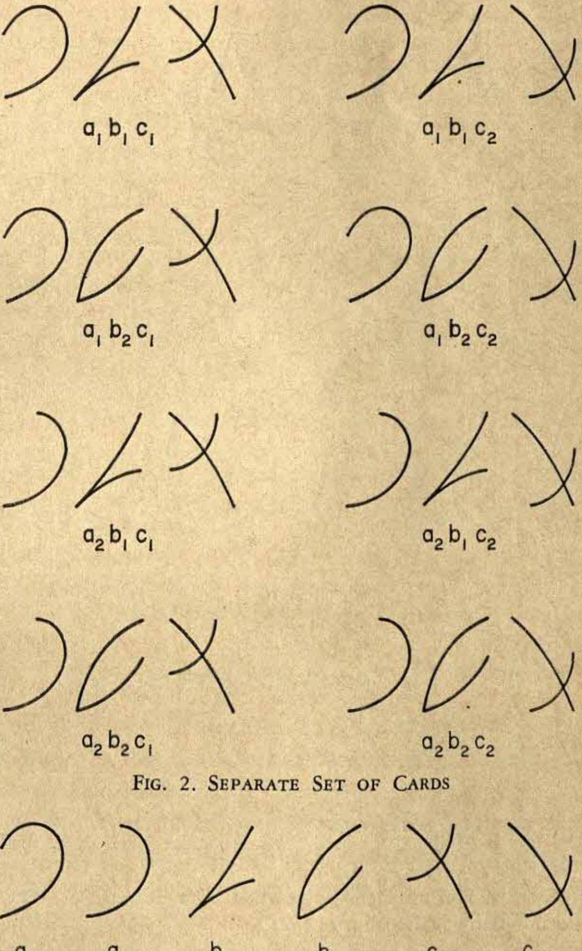


FIG. 2. SEPARATE SET OF CARDS

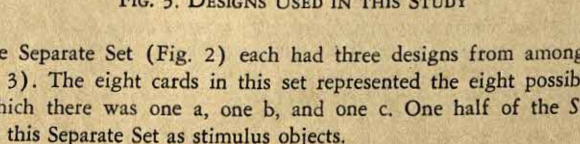


FIG. 3. DESIGNS USED IN THIS STUDY

cards in the Separate Set (Fig. 2) each had three designs from among designs a_1 to c_2 (Fig. 3). The eight cards in this set represented the eight possible combinations in which there was one a, one b, and one c. One half of the *Ss* were presented with this Separate Set as stimulus objects.

The other half were given the Combined Set, shown in Fig. 4. Each of the eight cards of the Combined Set contained the same three designs as corresponding cards in the Separate Set, but these were now so changed in position that the factor of smooth continuation favored their being perceived as a whole. Thus, the cards of the Separate Set were likely to be coded in units representing the separate designs a_1–c_2, while the cards of the Combined Set were more likely to be coded in units representing the entire patterns on each card.

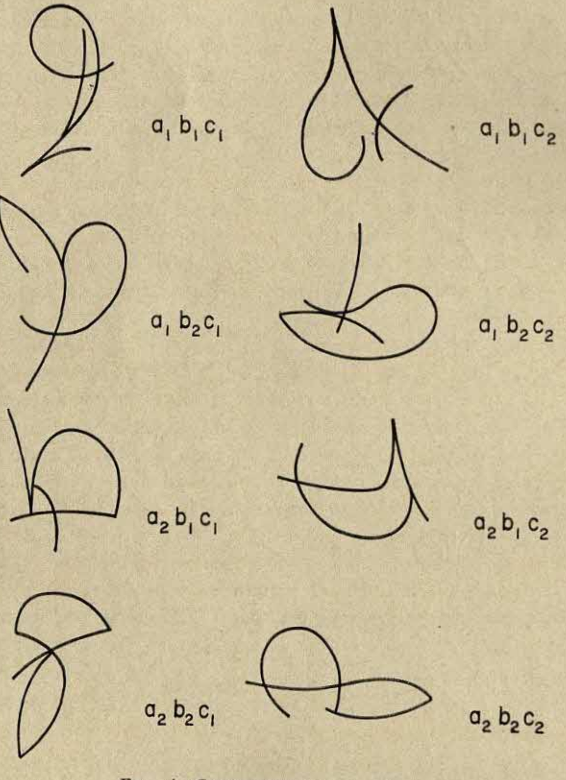


FIG. 4. COMBINED SET OF CARDS

Three different types of problems were used with different *Ss*. The cards were always divided into two equal groups of four.

In Type 1, the division was merely according to a, b, or c, alone. For example, one such classification grouped together the cards containing a_1 in one class and the cards containing a_2 in the other.

In Type 2, the division was made according to *two* designs—a and b, a and c, or b and c. For example, the cards containing either a_1 and b_1 or a_2 and b_2 were grouped together in one such classification, as against the cards containing either a_1 and b_2 or a_2 and b_1.

In Type 3, all three of the designs a, b, and c were relevant.

Whether the concepts in these types of problems differed in structure and number of cognitive units depended on how the stimuli were coded. If they were coded in units representing the separate designs a_1–c_2—as was likely with the Separate Set of cards—Type 1 was simplest and Type 3 most complex. If, however, the cards were coded in units representing the entire patterns on each card—as was more likely with the Combined Set—there would be no such difference, as the concept would be attained essentially by *S*'s memorizing which cards went into every class.[4]

The experimental hypotheses were thus as follows: (1) Over-all, concept-formation will be easiest with Problem Type 1, in which a, b, or c is relevant alone; and hardest with Type 3, in which all three are relevant; and (2) the variation in difficulty with different problem-types will be greater with cards of the Separate than with those of the Combined Set.

The *S*s (60 in number) were randomly divided into six equal groups in a 2 × 3 factorial design, with the type of problem and the set of cards as treatment-variables.

Every *S* was told that he would be shown eight cards, each one of which was called either 'X' or 'Y,' and asked to learn to label them correctly. Which cards were called 'X' and which 'Y' varied for different *S*s randomly within the type of problem to which *S* was assigned.

All eight cards were laid before the *S* (placement being randomly determined) and labelled in order. The experimenter called 'X' or 'Y' for a card every 15 sec. whenever *S* did not make a guess earlier. After the cards had been labelled, they were picked up and replaced in another order and the same procedure was followed again, until *S* had consecutively named all of them correctly. Then the cards were presented one at a time. The criterion of concept-attainment was correct labelling of all eight shown singly, twice in succession. The experiment was terminated if this did not occur within 90 min.

The difficulty of concept-attainment was measured by the number of cards presented until criterion was reached. (For those *S*s who did not reach the criterion within 90 min., the measure was "more than . . ."—whatever number had been presented.)

RESULTS AND DISCUSSION

Table I shows the median number of cards to criterion for each group. The significance of the over-all differences among problem types was computed by the Kruskal-Wallis technique; p was found to be smaller than 0.001.[5] Individual comparisons between pairs were then made, using the Mann-Whitney *U*-test.[6] Type 1 was found to be significantly easier ($p < 0.01$) than either of the other two types of problems, as predicted by Hypothesis 1. There was, however, no significant difference between Types 2 and 3. This may have been due to the fact that Type 2 with the cards of the Separate Set was already so difficult that 3 of the 10 *S*s did not reach criterion in the time allowed.

---

[4] That the Separate Cards indeed *were* categorized more frequently in units representing the separate designs, and the Combined Cards in units representing the whole patterns, was indicated by the *S*s' descriptions of their concepts at the end of the experiment.

[5] Sidney Siegel, *Nonparametric Statistics for the Behavioral Sciences*, 1956, 184-194.

[6] *Ibid.*, 116-127.

According to Hypothesis 2, the variation in difficulty should have been greater with the Separate Set of cards than with the Combined. The differences among the three problem-types within the Separate Set, again using the Kruskal-Wallis technique, were found to be significant at the 1/10% level of confidence, while the differences within the Combined Set were insignificant. Further, Type 1 with the cards of the Separate Set was significantly easier than any type with the cards of the Combined Set, while Type 3 with the Separate Set was significantly harder than any type with the Combined Set ($p < 0.05$).

These results lend some support to the interpretation of complexity as a function of structure and number of cognitive units, though these were not independently varied in the present experiment. The over-all differences in the difficulty of the different problem-types could, however, also be interpreted in terms of the likelihood of adoption of appropriate strategies. Bruner, Goodnow, and Austin have presented evidence that conjunctive strategies, at least in our culture, are the ones that people tend to prefer —and these strategies would be appropriate for Type 1 but not the other types.[7]

TABLE I

MEDIAN NUMBER OF CARDS TO CRITERION IN EACH GROUP

Type of problem	Set of cards		Both
	separate	combined	
1	7.5	65.5	8.5
2	149.5	80.5	89.5
3	376.0+	23.5	99.5
All	85.5	69.5	

The more indicative finding is the difference in the effect of problem-types with the different cards. Problem-type clearly made a difference with the cards of the Separate Set, while with the Combined Set it did not. This would seem to imply that it does not suffice to consider only objective attributes in concept-attainment, and that cognitive units indeed are relevant here.

SUMMARY

An experiment on concept-attainment was performed in which two different sets of cards—Separate and Combined—were given to different *S*s.

---

[7] Bruner, Goodnow, and Austin, *op. cit.,* 168 f.

The Separate Set consisted of cards with three separate designs on every card. Each card in the Combined Set contained the same three designs as one of the cards in the Separate Set, combined by a rule of smooth continuation.

In both groups, every *S* was given one of three different types of problems. In Type 1, only one of the designs was relevant to the concept; in Type 2, two were relevant in a contingent relation to each other; and in Type 3, all three were relevant—in the most complex relation.

There were significant differences in the difficulty of concept-attainment in the three types of problems, with Type 1 significantly easier than Types 2 and 3 with the cards of the Separate Set. With the Combined Set, there was no significant difference among problem-types.

The results support the view that the complexity of concept-attainment is a function of the structure of the concept and the number of cognitive units it involves. With the cards of the Separate Set, the *S*s tended to code each design in a different cognitive unit; hence, here problems of Types 2 and 3 involved a greater number of relevant units, in complex interrelations with each other, than in Type 1. With the Combined Set, the *S*s tended to code each *card* in a different cognitive unit; thus, here they simply memorized which cards were given which label, no matter what the problem type.

INCIDENTAL AND INTENTIONAL LEARNING UNDER THREE CONDITIONS OF MOTIVATION

By RONALD JOHNSON and CALVIN THOMSON, San Jose State College

McGeoch and Irion define incidental learning as "learning which apparently takes place without a specific motive or a specific formal instruction and set to learn the activity or material in questtion."[1] In this definition, the operational meaning of incidental learning depends upon the adequacy of control that has been exercised over the set to learn. A study by Bahrick, Rankin, and Fitts, and one by Postman and Senders, have shown that *S*s use self-instruction when learning in situations in which there is neither apparent motive nor specific formal instruction.[2] Bahrick attempted to control self-instruction in incidental learning by increasing the incentive associated with the learning of the intentional material.[3] He found a significant decrease in the amount of incidental learning when the motivation to learn the intentional material was increased.

The present study was planned to determine the effect of various levels of motivation on the rate of incidental and intentional learning. It differed from similar studies in one respect. The incidental material to be learned was not neutral with regard to the intentional task, but, instead, constituted a constant distracting or interfering influence on intentional learning. This 'noxious' quality seemed likely to reduce the possibility of self-instructed sets to learn the incidental material.

METHOD

Learning materials. The materials to be learned consisted of two lists of nonsense-syllables, 10 syllables in each list, each syllable of 93%–100% association value by Glaze's norms.[4] The two lists were entirely different, with no syllables in common. Since the learning-task required a vocal response, each syllable was selected to sound dissimilar to the others.

---

* Received for publication November 4, 1959.

[1] J. A. McGeoch, and A. L. Irion, *The Psychology of Human Learning*, 1952, 210.

[2] H. P. Bahrick, R. E. Rankin, and P. M. Fitts, Effect of incentives upon reaction to peripheral stimuli, *J. exp. Psychol.*, 44, 1952, 400-406; Leo Postman and V. L. Senders, Incidental learning and generality of set, *ibid.*, 36, 1946, 153-165; J. G. Jenkins, Instruction as a factor in incidental learning, this JOURNAL, 45, 1933, 471-477.

[3] Bahrick, Incidental learning under two incentive conditions, *J. exp. Psychol.*, 47, 1954, 170-172.

[4] J. A. Glaze, The association value of nonsense syllables, *J. genet. Psychol.*, 35, 1928, 255-269.

The task. The *Ss* worked in pairs, of which both members were drawn from the group having the same motivation. The two *Ss* were seated side by side, each before a memory-drum which contained one of the two totally dissimilar lists of nonsense-syllables. Their learning-trials were simultaneous, with each *S* required to anticipate verbally in a "good, loud and clear" voice, each syllable on the list. Each syllable appeared for 2-sec. There was a 2-sec. interval between successive syllables and an 8-sec. interval between successive trials. All *Ss* were given 25 trials. *E* stood behind each *S* to record responses and the time-intervals between trials. The intentional task was to learn the prescribed list; the incidental task was to learn the syllables from the partner's totally dissimilar list. The *Ss* received no instructions concerning the list learned by the partner; their instructions were merely to learn their own lists as rapidly as possible. Immediately following the twenty-fifth trial, each *S* was handed a piece of paper and instructed to write down the syllables he had learned. Each was then presented with a typed list of 30 nonsense-syllables (of the same Glaze associative value) within which appeared, in random order, the 10 syllables on the partner's list. He was asked to circle the 10 syllables learned by his partner and urged to guess, if he was not certain. This test of recognition constituted the measure of incidental learning.

Subjects. The *Ss* were 30 freshmen from introductory psychology. Ten of the *Ss*, designated as high motivation (*HM*), were selected from one of the author's classes. Five of these (*HM*) *Ss* had received course grades that were on the borderline between C and D, and five were on the borderline between B and C in course grades. To develop motivation for this task, an agreement was made with each *S*. If he learned the experimental material more rapidly than the others in his group (C-D or else B-C) and if, after the final class examination, his grade was still borderline, he would receive the higher of the two possible grades. Thus, the *Ss* in this group had one chance in five of winning special consideration relative to final course grades. The medium motivation (*MM*) group were led to believe that the rate of intentional learning of nonsense-syllables was a valid and reliable measure of intellectual ability. The low motivation (*LM*) *Ss* were selected from other class sections and were told that their performance was for the purpose of helping out in an experiment and had no relation to the course or course grades. The average course grade received by members of each group was approximately equal.

RESULTS AND DISCUSSION

Table I gives the mean number of correct anticipations made on 25 trials by members of the three experimental groups for each syllable and, in addition, the mean for the entire list.

The *MM* group performed significantly ($t = 3.06$, $P = < 0.01$) better on the intentional task than did the *HM* group. The differences between the *MM* and *LM* ($t = 0.73$, $P = 0.5$ to 0.4) and the *LM* and *HM* ($t = 1.88$, $P = 0.1$ to 0.05) groups are not significant. There is a greater variability in number of correct anticipations among the *LM* and the *HM* groups as compared with the *MM* groups. This difference approaches significance (*LM* vs. *MM*, $F = 2.05$, $P = < 0.06$). It suggests that a medium level of motivation is both most effective and most uniform in its effect on rate of intentional learning.

Incidental learning was measured by the mean number of correct recognitions of syllables from the partner's list. Since each *S* was required to select 10 syllables out of 30, one would expect 3⅓ correct recognitions by chance. The actual mean recognitive scores for the three groups are presented in Table II.

The *LM* and the *MM*, but not the *HM*, groups showed significant amounts of incidental learning, even when the incidental material was 'noxious' (*i.e.* interfering and distracting) with regard to intentional learning. The likelihood of *S*s' having self-instructed sets to learn the incidental material seems remote. Hence, this finding offers more support than do previous studies to the position that learning without set *can* occur. The amount of motivation for the intentional task seems an important determiner of whether learning without set *does* occur in any given situation. This last finding may have relevance in resolving the apparent conflicts between the results of various latent learning studies.

TABLE I

Mean Correct Anticipations of Material Intentionally Learned in Three Groups Differing in Motivation

N = 10 for each group. All *S*s received 25 trials.

Motivation	Syllable										Total list	*SD*
	1	2	3	4	5	6	7	8	9	10		
Low	21.9	20.8	18.5	19.0	15.4	17.3	17.8	17.0	18.8	21.0	18.8	3.7
Medium	23.7	22.8	21.4	19.8	15.9	18.3	18.0	18.4	20.2	21.3	20.0	1.8
High	21.2	17.9	17.5	13.0	9.8	12.5	13.2	14.1	16.3	19.2	15.7	3.6

TABLE II

Mean Recognition Scores of Material Incidentally Learned by Three Groups Differing in Motivation

N = 10 for each group. Chance score is 3.3 correct.

Motivation	Number correct	*SD*	*p* that difference is chance			
			Chance	*LM*	*MM*	*HM*
Low	5.4	1.25	<.001	—	.4	.02
Medium	4.5	1.40	.011	.4	—	.2
High	4.0	1.40	.14	.02	.2	—

The motivational groups vary in the mean amounts of intentional as well as incidental learning. There is no significant difference between the groups in mean numbers of correct anticipations of Syllables 1, 2, 3, 8, 9, and 10. The existing differences in rates of intentional learning are made up of the differences in the frequency of correct recognition of Syllables

4, 5, 6, and 7. The differences are found only among those syllables that would be most affected by the differing levels of interference produced by differing levels of motivation. Hence, the obtained differences between the motivation groups seems likely to be due to differences in motivation rather than differences in learning ability.

The problem of whether differences in incidental learning are due to differences in exposure to the incidental material (the partner's intentional learning task) or to differences in motivation must still be resolved. If frequency of exposure to incidental materials as measured by the partner's correct anticipations, is related to the amount of incidental learning shown by an *S*, then there should be a significant positive correlation between the amount of a *S*'s incidental learning and the number of correct anticipations made by the partner. The actual correlations (r) between these two variables obtained in the three motivation groups are: $LM = -0.36$; $MM = +0.18$; $HM = -0.18$. None of these correlations is significant. It appears that frequency of exposure to incidental materials is not related

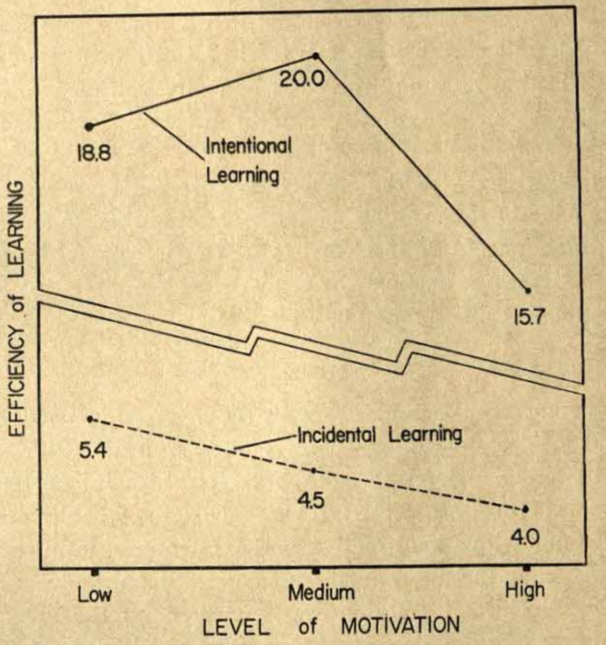


FIG. 1. RELATIONSHIP BETWEEN LEVELS OF MOTIVATION AND MEASURES OF LEARNING INTENTIONAL AND INCIDENTAL MATERIAL

to the amount of incidental learning shown by the *S*s in this study. The pattern of significant differences found between the groups in amounts of intentional and incidental learning also seems to suggest that frequency of exposure is not the crucial factor in incidental learning.

Fig. 1 provides a graphic illustration of the effect of varying levels of motivation on the two types of learning. It indicates that motivation affects the speed of intentional and of incidental acquisition differentially. As has frequently been claimed, motivation apparently has a curvilinear effect on intentional learning, with a medium level of motivation being most effective in facilitating learning. The effect of motivation on incidental learning seems to be essentially linear, with the most incidental learning taking place under the lowest level of motivation.

This finding seems, to the authors, to raise questions regarding what the optimal level of motivation should be, in order best to facilitate learning. If a high amount of incidental as well as intentional learning is desirable in a learning situation, the results would indicate that low, rather than medium, motivation might be desirable.

SUMMARY

The experiment was designed to study the effects of motivation on intentional and incidental learning. Three groups of introductory psychology students, differing in level of motivation, learned a list of 10 nonsense-syllables. The *S*s worked in pairs, responding to the syllables presented on a memory-drum with a loud vocal response. Intentional learning was measured by obtaining the mean number of correct anticipations at each syllable-position. Incidental learning was measured by the frequency that syllables learned by the *S*'s partner were recognized by the *S*. The results indicated that (1) incidental learning occurred even when the incidental material was distracting rather than neutral with regard to the intentional learning task, and (2) motivation appeared to have a differential effect on incidental as opposed to intentional learning. A medium level of motivation facilitated intentional learning; a low level of motivation was best for incidental learning.

THE ROLE OF SLANT IN THE PERCEPTION OF SHAPE

By HANS WALLACH and MARY E. MOORE, Swarthmore College

When a figure or the plane surface of an object is oriented obliquely to the line of sight, the shape of its retinal image differs from its real shape; if the perceived shape nevertheless approximates the objective shape, we call this an example of 'shape-constancy.' The traditional explanation of shape-constancy assumes that the orientation of the figure or surface is registered, and its slant relative to the line of sight taken into account, as the shape is perceived. This explanation is derived by analogy to other constancies, as, for example, size-constancy, where distance is taken into account. There is, however, no empirical evidence in the literature that the perceptual process follows such a course. Experiments in the area hardly go beyond the demonstration of shape-constancy. As a matter of fact, that the perception of shape depends on cues for slant is not even a logically necessary assumption.

To talk separately of slant and shape makes little sense where, with monocular observation, only perspective cues for slant are given. Such perspective cues consist in the deformation of the retinal images of familiar shapes and they imply 'assumptions' about the true shapes. Only when these assumptions are correct, that is, when the shapes viewed are in fact the familiar ones, will adequate cues for slant exist, but in that case, cues for slant are no longer needed; the 'assumptions' themselves can account for the perception of familiar shape.

When a slanted surface is binocularly observed and stereoscopic vision operates, the notion that both shape and slant result directly from the given disparities is quite feasible. Those who think in terms of an inherently three-dimensional perceived space with an inherent geometry would take such an assumption for granted. When, for example, the two oblongs of Fig. 1 are stereoscopically combined, the given disparity causes the right vertical edge of the perceived rectangle to be located in a frontal plane a certain distance behind the frontal plane of the left edge. Hand in hand with this oblique position goes a greater width of the perceived oblong. When the existence of a perceptual space is assumed in which the

---

* Received for publication December 9, 1960. This work was supported by a grant from the National Science Foundation.

primary stereoscopic effect takes place, the geometry of the space, whether it be innate or established by learning, will determine the greater distance between the vertical edges of the obliquely oriented oblong. Slant and shape will be equally direct results of the stereoscopically caused depth between the oblong's vertical edges.[1]

These considerations, of course, show merely that the postulate of a process which takes tilt into account may be unnecessary; they do not preclude the existence of such a process. We can, therefore, be sure of the facts only by setting up an experiment in which such a process, if it existed, could assert itself. This can be done by creating conditions under which cues for slant are more adequate than cues for shape. If these conditions yield constancy commensurate with the cues for slant, the potential of such cues would be demonstrated.[2]

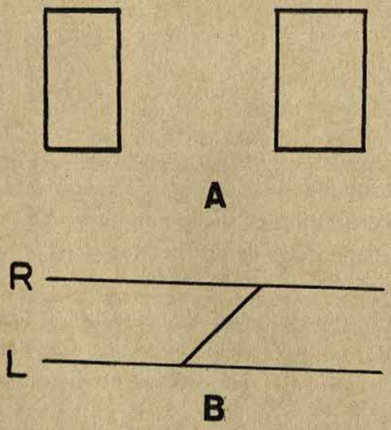


FIG. 1. DIAGRAMS SHOWING: (A) STEREOGRAM OF AN OBLONG STANDING OBLIQUELY IN SPACE; AND (B) GROUND-PLAN OF THE PERCEIVED ARRANGEMENT IN DEPTH

L is the frontal parallel plane of left edge, *R* the frontal parallel plane of right edge.

Such an experiment was performed with black isosceles triangles as the slanted figures. A triangle was chosen as the stimulus because it does not convey any perspective cues for slant when the tip is tilted back from the frontal parallel plane. The triangles were fastened to a slanting white surface of oblong shape whose edges were clearly visible against a black table top. An arrangement was used which permitted *O* to see the edges of the oblong surface with both eyes but restricted his view of the triangle to one eye only. The slant of the oblong background, then, was represented by stereoscopic as well as perspective cues, while the figure itself was stripped of both of these, the major cues for shape and slant. Inasmuch as the triangle usually is seen to lie in the plane of the white surface, the cues for slant of the oblong background also could serve as cues for the slant of the triangle. The results to be reported below show that this was indeed the case.

---

[1] For different reasons, Gibson, too, would hold this view. He would expect 'shape-at-a-slant' to be the simple and immediate result of stimulation.

[2] This plan differs from that of B. K. Stavrianos (*Arch. Psychol.*, 49, 1945, No. 296, 1-94) and of Jacob Beck and J. J. Gibson (*J. exp. Psychol.*, 50, 1955, 125-133). These authors were concerned with the question whether apparent shape and apparent slant are linked. We are asking whether *cues* for slant can have an effect on apparent shape.

Method. Triangles, with a base of 2.5 in., were presented on two slopes. Under Condition 1, a 4-in. high triangle rested on the oblong surface when its slant was 59.5°, that is, when it formed an angle of 59.5° with the frontal plane of *O*. Under Condition 2, this tilt was 46.5° and the triangle had an altitude of 2.95 in. The two conditions were so chosen that the triangles produced retinal images of identical shape. The perceived shape of the triangles was measured by a matching procedure. A series of comparison-triangles was used ranging from 1.6–4.4 in. in altitude in steps of 0.1 in. Like the two standard triangles, they were made of black cardboard, were isosceles, and had a base of 2.5 in. Each was fastened to a white cardboard 10 × 10 in., which could be inserted in a wooden frame standing in frontal parallel position 12 in. from *O*'s eyes. Beyond the frame was the slanting white surface on which the standard triangle was presented, its center 31 in. from *O*'s eyes. By pushing a cardboard with a comparison triangle into the frame and pulling it halfway out, *E* could give *O* alternate views of the standard triangle and of a comparison-triangle. The position of the comparison-triangle, when fully inserted in the frame, was such that it was visible binocularly to *O*, while the small obstruction in front of *O*'s left eye blocked out the standard triangle for this eye, but not the edge of the white surface on which it rested. A head-rest served to maintain these relations throughout the experiment. An abbreviated method of limits was used for the matching, with two ascending and two descending series presented in alternation. To give *O*s practice in matching, a third standard triangle, smaller than either of the experimental triangles, was matched initially, using one ascending and one descending series. Sixteen *O*s were studied in this experiment.

In addition to the main experiment, two control experiments were performed. In both, cues for the slant of the background surface on which the standard triangle rested were eliminated by inserting a screen with a circular hole between the frame for the comparison-triangles and the display for the standard figure. The size of the hole was so chosen that only the triangle and part of the homogeneous white surface on which it rested were visible, but not the edge of this surface. In the first of the control experiments, *O* wore an eye-patch and had to observe monocularly; in the second, he was allowed to see the triangles with both eyes. Eight *O*s participated in the monocular and 16 in the binocular control experiment. The measuring procedure was the same as in the main experiment.

Results. The matches in the main experiment show a high degree of constancy. The mean perceived altitude for the 4-in. triangle was 3.4 in. when this triangle was slanted to produce a projection on the frontal plane of 2.03 in. in height. For the other standard, whose altitude was 2.95 in., and which, because of smaller slant, produced the same projection, the mean altitude match was 2.71 in. The difference between these means is highly reliable ($t = 12.62$, $p < 0.001$). Since the triangles produce identical retinal images, the only difference in stimulation which can account for the difference in the perceived altitude of the triangles is in the cues for their slant.

In the first control experiment, the monocular view through the screen was supposed to represent reduction-conditions. Thus, by comparing the results of this experiment with those of the main experiment, we can obtain the traditional proof of constancy, a difference between matches under reduction- and under constancy-conditions. In the first control experiment, the mean perceived altitude of the two standard triangles was small and nearly the same; namely, 2.41 and 2.28.[3] The differences between these values and the means of the altitudes perceived in the main experiments reported above were highly significant as the difference for the 4-in. triangle showed a t of 15.12 ($p < 0.001$) and, for the 2.95-in. triangle, t equaled 8.52 ($p < 0.001$). For both standard triangles, the matches obtained in the main experiment were close to the true shapes of the triangles and those obtained in the monocular control experiment were close to the shape of their projection on the frontal plane (Table I). In the main experiment, therefore, the slant-cues for the surface on which the monocularly viewed standard triangles rested produced a strong constancy-effect, and there can be no doubt that cues for slant influence perceived shape.

TABLE I

MEAN PERCEIVED ALTITUDE OF TRIANGLES (IN IN.)

Objective altitude	Projective altitude	Main experiment	First control	Second control
4.00	2.03	3.40	2.41	3.58
2.95	2.03	2.71	2.28	2.83

Just how complete is the shape-constancy thus achieved? The answer comes from a comparison between the matches obtained in the main experiment with those of the second (binocular)control experiment where conditions for shape-constancy are optimal. As Table I shows, the binocular matches are not very different from those obtained in the main experiment. This difference is barely significant for the 4-in. triangle ($t = 2.41$; $p < 0.05$), but not for the 2.95-in. triangle ($t = 1.98$; $p > 0.05$). The binocular match of 3.58 in. for the 4-in. triangle whose slant was almost 60° implies a sizeable lag in constancy, which is in keeping with the results usually obtained for large slants.

---

[3] The difference of 0.13 in. proved, however, to be significant. No attempt was made to find the reason for the difference.

Conclusion. In this investigation, good shape-constancy for a figure was obtained solely from cues to the slant of its background. To be sure, the investigation dealt with an entirely artificial situation. Its significance lies in the fact that it provides a clear instance where, in the perception of shape, slant *per se* is taken into account, and that the existence of a process in which perception of shape takes this route is thus demonstrated. No definite claim can be made for the importance of this kind of process in everyday perceptual functioning. As was shown above, there are no compelling reasons for postulating the occurrence of such a process under ordinary conditions. On the other hand, this does not mean that, once the existence of such a process is demonstrated, it is unlikely to participate in ordinary perception of the shape of slanting surfaces. In three-dimensional form-perception, duplication of function is the rule rather than the exception. The importance of the demonstration lies in its theoretical implications. One may now, for instance, take the view that stereoscopic perception directly determines only depth and slant, but that the shape of the slanted surface is the product of a process by which slant is taken into account.

THE RATE OF LEARNING A DISCRIMINATION IN RELATION TO SOME PREDICTIVE MEASURES

By EDWARD J. GREEN and ROGER W. SQUIER, JR., Dartmouth College

It has been demonstrated that changes in frequency of responding may be predicted by simple statistical models of learning. That is, the amount of change in the rate of operant responding is dependent upon the probabilities that the organism will sample, discriminate, or perceive various stimulus elements available to the organism on a given occasion.[1] The probability of sampling so defined has been labelled the *sampling* ratio.[2] This ratio is further defined as 's/S,' where 's' is the number of elements sampled by a given organism, and 'S' is the set of elements available to be sampled. It has been demonstrated that changes in *E* do in fact alter the rate of change in conditioning. Pribram has found that surgical interference in certain areas of the temporal cortex produces changes in rate of conditioning that would be predicted by a reduction in 's.'[3] The present study was designed to relate performance commonly associated with intellectual ability to speed of conditioning in a standard discriminative problem. It follows that slow learners are likely to be low in rank as they are measured by standard tests of intelligence. The present experiment was designed specifically to test two implications of statistical theories of discriminative learning.[4] The first is that 'bright' *S*s should show more rapid acceleration in discriminative learning than 'dull' *S*s. The second prediction has to do with variability in responding. According to the concept of sampling ratio, *S*s with small ratios should be more variable throughout their performance than *S*s with large sampling ratios, *i.e.* rapid learners will show a smoother, more consistent performance than will slow learners.

Method. Eighteen *S*s were chosen from a class in introductory psychology on the basis of their ranks on the College Predictive Index (*PI*). The *PI* is composed of

---

* Received for publication July 21, 1959.

[1] E. J. Green, Stimulus-variability and operant discrimination in human subjects, this JOURNAL, 69, 1956, 269-273.

[2] W. K. Estes, Toward a statistical theory of learning, *Psychol. Rev.*, 57, 1950, 94-107.

[3] Karl Pribram, personal communication.

[4] E. J. Green, A simplified model for stimulus discrimination, *Psychol. Rev.*, 65, 1958, 56-63.

appropriate weightings of the CEEF SAT Verbal Score, the SAT Mathematical Score, the Headmaster's or Principal's rating, and the secondary school class ranking. Nine of the *S*s had *PI*s which were in the ninety-seventh percentile or above. These composed the 'high' group. Nine Ss, the 'low' group, had *PI*s in the sixth percentile and below. *S*s were placed in the experimental room singly for an hour's experimental session. A different visual stimulus-pattern was presented continuously on a panel of 10 jewel lights throughout the session. One light was always associated with reinforcement of the response, five were never so associated, and two lights were associated with reinforcement exactly half the time. During the intervals selected for reinforcement, *S* received points on an electrical counter visible to him. The points were awarded on a 10:1 variable ratio-schedule for pulling a hand-lever. That is, every tenth pull, on the average, resullted in an increment in the *S*'s score. *S* was instructed to get as high a score as possible, find out the rule that made a pattern an occasion for getting points, and to respond only when he thought a pattern was an occasion for reinforcement. Illumination of the room was restricted to that provided by the stimulus-panel, and a small blue bulb mounted over the counter. White noise introduced both through headphones and through a speaker in the room was used to mask extraneous sounds.

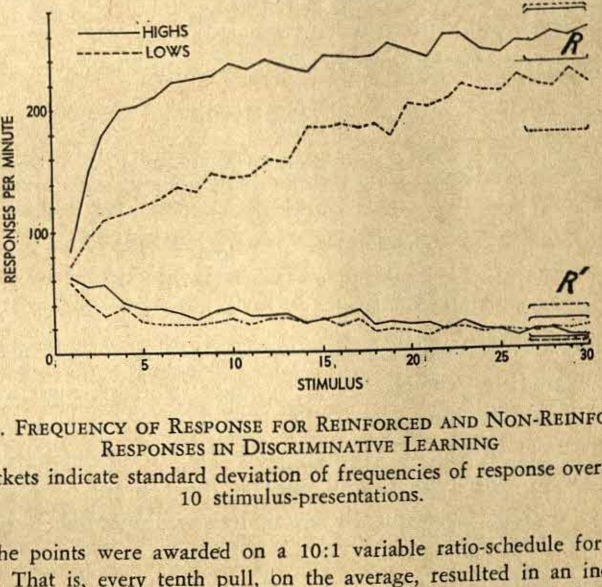


FIG. 1. FREQUENCY OF RESPONSE FOR REINFORCED AND NON-REINFORCED RESPONSES IN DISCRIMINATIVE LEARNING

Brackets indicate standard deviation of frequencies of response over last 10 stimulus-presentations.

Results. As seen in Fig. 1, the 'high' *S*s did accelerate more rapidly than did the 'low' *S*s in the reinforced (*R*) condition. There did not appear to be any difference between the groups during the nonreinforced (*R′*) condition. It should be noted, however, that since the initial rate of response to *R′* by the high *S*s was somewhat above that of the low *S*s, the high *S*s had farther to drop, and did go somewhat below the terminal rate of the low *S*s by the end of an hour; so while the differences between the groups in responding to *R′* are not statistically significant, they are in the predicted direction.

Of greater significance is the analysis of the variability between groups. Variance was computed for the rate of responding over the last 10 stimulus-presentations in the presence of both *R* and *R′* for the two groups. The results of the preliminary analysis are shown in Fig. 1. Brackets indicate relative magnitudes of standard deviations appropriate to the four conditions. From this analysis, the prediction of greater variability in the responding of the low *S*s is amply supported by *F*-ratios significant beyond the 0.001 level. But two problems emerge. First, the overall learning curves are not asymptotic in either *R* or *R′*; and *R*-curve of the high *S*s in particular has a slope nearer zero than the low *S*s. The conditioning process contaminates, therefore, the variances in a direction favoring the hypothesis. Secondly, inter-*S* variability is considerable in both groups, and absolute changes in rate of response undoubtedly do not have the same meaning for *S*s with high overall rates as they do for *S* of low overall rates of responding. Consequently, an analysis was undertaken to eliminate both sources of contamination.

TABLE I

VARIANCES COMPUTED ON THE BASIS OF PERCENTAGE DEVIATIONS OF RESPONSE-FREQUENCIES FROM INDIVIDUAL REGRESSION LINES

Condition	Group		*F*
	high	low	
R	8.94	9.11	1.019
R′	32.29	32.12	1.005

The mean rate of change over the last 10 stimulus-presentations to both *R* and *R′* stimuli were computed for each of the 18 *S*s. With this value for a slope a straight line was fitted to the data for each *S*. Deviations were then taken from this fitted line and were converted to percentages of the theoretical values. These percentages were squared, summed, and divided by the appropriate *N* to yield the variances presented in Table I. No *F*-ratio achieves significance between groups, since the variances are about as homogeneous as possible. While the actual group-curves are not straight lines, they are near enough to being straight that a linear regression-line introduces no serious errors.

Discussion. The first prediction, concerning rate of learning as related to the Predictive Index (*PI*), is verified. The second prediction, however, which would be an independent verification of the sampling-ratio concept, was found not to be supported by the data. In the free operant situation employed, no differences whatsoever were found in response-variability between groups showing high rate of change in conditioning and groups showing a low rate of change. The methodological point must be emphasized that had the analysis of the data been terminated at the level of raw frequencies of response, the prediction would have been verified in detail; but that such verification would have been spurious.

Of present interest is the possibility that the free operant discrimination might be developed as a diagnostic aid in practical problems. The technique has certainly proven to be sensitive to differences in ability in a college population with an attenuated range. Undoubtedly application to more general diagnostic problems is possible.

SUMMARY

On the basis of the statistical theory of learning, it was predicted that an index of academic ability would be correlated with performance in an operant discriminative task. The results supported the prediction that *S*s scoring low on the Prediction Index (PI) should be slow in developing a discrimination. A second prediction that intra-*S* variability in operant responding should be greater for low levels on the *PI* was not, however, supported.

The free operant discrimination is sensitive to differences among students measured by predictors of academic success and, it is suggested, may have practical uses as a diagnostic tool.

EFFECT OF STIMULUS-DURATION ON AUDITORY REACTION-TIME

By DAVID H. RAAB, Brooklyn College

The loudness of a brief burst of noise depends on its *energy*. Miller has shown that, at threshold and above, there is temporal integration of acoustic power up to around 0.1 sec.[1] The reciprocity of duration and intensity is somewhat less than perfect (8.8 db. per log unit of time), and the critical duration is itself inversely related to intensity.

In contrast to this 'long term' effect is the short time integration reported by Hall.[2] Using the procedure of binaural fusion and centering. Hall found the apparent intensity of a noise burst to increase with burst duration up to around 10 m.sec.[3] Tobias and Schubert, who also employed dichotic stimulation, have calculated the 'onset duration' of a burst to be of the order of 3 m.sec.[4]

The present study of burst durations utilizes a third measure of 'stimulus strength,' viz. simple reaction-time (*RT*). Since loudness grows with stimulus-duration up to a time which is not much shorter than the auditory *RT*, it seems reasonable to suppose that the latter depends on a briefer time-integral of power than the former. In a previous report from our laboratory, the visual *RT* was found to be unchanged by shortening the flash-duration from 500 to 10 m.sec.[5] For luminances between 0.3 and 3000 ft.-L., foveal *RT* depends on a very brief 'package' of luminous energy, although brightness grows with duration for at least 0.5 sec. Comparable

---

* Received for publication December 19, 1960. This study was supported by Grant B-1028 from the National Institute of Neurological Diseases and Blindness of the U. S. Public Health Service and by funds provided by Brooklyn College. The data were collected by Miss Maxine Lanskrow and Miss Naomi Maizel.

[1] G. A. Miller, The perception of short bursts of noise, *J. acoust. Soc. Amer.*, 20, 1948, 160-170.

[2] J. L. Hall, A psychoacoustic study of the mechanism of binaural fusion, Unpublished M.S. thesis, Massachusetts Institute of Technology, 1959.

[3] See B. H. Deatherage and I. J. Hirsh, Auditory localization of clicks, *J. acoust. Soc. Amer.*, 31, 1959, 486-492; E. E. David, Neuman Guttman, and W. A. van Bergeijk, Binaural interaction of high-frequency complex stimuli, *ibid.*, 31, 1959, 774-782.

[4] J. V. Tobias and E. D. Schubert, Effective onset duration of auditory stimuli, *ibid.*, 31, 1959, 1595-1605.

[5] D. H. Raab, E. Fehrer, and M. Hershenson, Visual reaction time and the Broca-Sulzer phenomenon, *J. exp. Psychol.*, 61, 1961, 193-199.

data are lacking for the auditory system. In the studies by Brogden, *et al.*,[6] on response-terminated tones, the shortest fixed duration used was 100 m.sec.

Method: Apparatus. The noise stimuli were generated by a Grason-Stadler 455B noise generator, and band-limited (100–7000 ~) by a Kronhite 330-M filter. Bursts were produced by means of a Grason-Stadler 829S electronic switch which was triggered by a set of Tektronix Pulse and Wave-form Generators. The stimuli were attenuated and delivered to a PDR-10 earphone mounted in a doughnut cushion.

A Berkeley 7350 EPUT counter was used to measure *RT*s, and to calibrate the durations generated by the Tekronix units.

Procedure. Six burst-durations (2, 5, 10, 20, 50, and 100 m.sec.) and two sensation-levels (40 and 60 db.—determined with duration equal to 1.0 sec.) were employed. These stimuli were delivered through one earphone. *RT*-trials were presented once every 12 sec. Each trial bgan with a 1000 ~ warning tone, lasting 0.75 sec., delivered to the opposite earphone. The noise-stimulus was presented (and the *RT*-counter started) 1.75, 2.00, or 2.25 sec. after the end of the warning tone. The counter was stopped when *S* depressed a normally-closed telegraph key.

TABLE I

MEAN REACTION-TIME AS A FUNCTION OF STIMULUS-DURATION

Sensation level	*S*	Noise burst-duration (m.sec.)					
		2	5	10	20	50	100
40 db.	*ML*	144.5	142.1	137.8	135.5	132.0	131.5
	NM	133.6	128.6	125.3	123.7	123.4	121.1
	DR	127.3	124.9	122.5	118.3	114.7	117.0
	AS	124.8	123.7	122.9	114.1	108.9	109.1
	Mean	132.6	129.8	127.1	122.9	119.7	119.7
60 db.	*ML*	137.5	137.2	134.0	128.4	128.1	127.2
	NM	126.8	122.4	123.1	119.4	118.9	119.1
	DR	121.5	119.5	115.2	113.6	112.2	113.1
	AS	124.9	119.3	114.8	110.8	107.3	108.5
	Mean	127.7	124.6	121.8	118.0	116.3	117.0

Blocks of 18 trials were separated by 1-min. rest-periods. Three such blocks (54 trials), preceded by four practice-trials, constituted a test-session. Two sessions (108 trials), separated by 15 min. of rest, were run each test-day.

Foreperiods (3), durations (6), and intensities (2) were randomized over trials. Each of the 12 intensity-duration pairs was presented nine times during a test-day. The longest and the shortest of the nine *RT*s were discarded, and the mean of the remaining seven was computed. Testing over 12 days thus yielded means based on 84 trials.

Subjects. Three undergraduate students and the author served as the *S*s. All were well practiced in the task before data were recorded.

---

[6] See L. W. Gregg and W. J. Brogden, The relation between reaction time and the duration of the auditory stimulus, *J. comp. physiol. Psychol.*, 43, 1950, 389-395.

Results and discussion. The results for each *S* are presented in Table I. As noted above, individual means are based on 84 trials. Standard deviations for these means were also computed. They are all of the order of 15 m.sec., and include inter-trial variations within a test-day as well as day-to-day variability. Data averaged for the four *S*s are plotted in Fig. 1. Each point is the mean of 336 measurements.

The *RT* to a 60-db. noise-burst decreases as burst-duration increases up to around 20 m.sec. Further increase in duration has no effect on the *RT*, although loudness would continue to grow with duration up to 65 m.sec.[7]

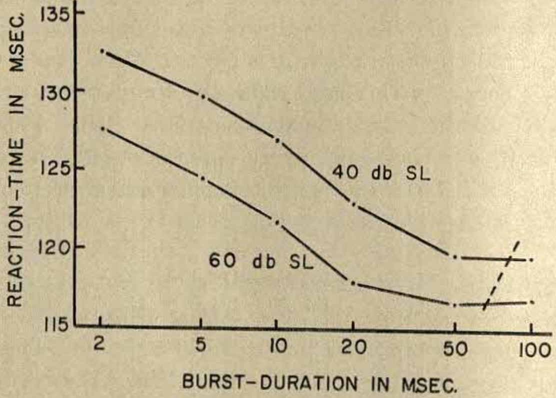


FIG. 1. REACTION-TIME AS A FUNCTION OF STIMULUS-DURATION
Each point is the mean of 336 reactions. The loudness of the bursts grow with duration up to the time indicated by the dashed line which intersects the two functions.

The 40-db. stimulus yields a minimal *RT* by 50 m.sec.; its loudness would grow with duration for 80 m.sec.

It is clear that temporal integration of power proceeds longer in the generation of loudness than in the mediation of the *RT*. As is true for the visual system, the critical duration is longer when *S*'s task is to estimate sensory magnitude than when his *RT* is measured.[8] For the data reported here, the two durations stand roughly in the ratio of 2:1. For the visual system, the ratio is probably much larger.[9] In any case, it is clear that there is no single '$I \times t$ law' for the visual or for the auditory system.

---

[7] Miller, *op. cit.*, 162.

[8] Raab, Fehrer, and Hershenson, *op. cit.*, 198.

[9] We do not know why this is so, but we are now extending the previous results on the Broca-Sulzer phenomenon to include durations shorter than 10 m. sec.

Of especial interest is the finding that the 2-m.sec., 60 db. and the 10-m.sec., 40-db. bursts yielded the same *RT*. The equivalence of 20 db. of power and only 7 db. (5×) of time was unexpected. The mean *RT* to the 40-db., 2 m.sec. stimulus was 132.6 m.sec. It was shortened equally by a hundred-fold increase in power or by a five-fold increase in duration.

The four *S*s were tested again in a second experiment. *All agreed that the 60-db. burst was the louder.* As an over-all check on the relative energy per burst for bursts of different power and duration, the output of a microphone placed near the earphone was measured with a Ballantine 320 RMS voltmeter. The meter's indication showed the expected relations between energy, amplitude and duration of burst. Four hundred additional *RT*-trials were then run, using only the two stimuli in question. Mean *RT*s to the 40-db. and 60 db. bursts were respectively 123.5 and 124.1 m.sec., *i.e.* virtually identical.[10] We note that the superiority of time over intensity is preserved until the critical duration is reached, after which duration ceases to affect *S*'s response. In the generation of the *RT*, stimulus-energy appears to be utilized in accord with a temporal-weighting function which favors energy delivered later.

SUMMARY

The simple reaction-time (*RT*) to a noise burst was measured as a function of burst-duration. Two intensities of noise were employed, and duration was varied between 2 and 100 m.sec. The *RT* decreased with increasing duration, reaching a minimum in approximately half the duration required for the full growth of loudness. Stimulus-energy delivered at the start of this critical duration is less effective than energy delivered toward the end. The results of this and related studies are taken to mean that there is no one 'Reciprocity Law' for a sensory system. The trading function which relates intensity and duration depends on the performance being tested.

---

[10] The means for this experiment, involving only two stimuli, are shorter than the *RT*s which appear in Table I. The difference may be due to increased practice or to some sort of context-effect.

THE EFFECT OF PRE-INSPECTION CONTROL MEASURES ON THE SIZE OF KINESTHETIC AFTER-EFFECTS

By PAUL BAKAN and RICHARD THOMPSON,
Michigan State University

In their original studies of kinesthetic after-effects, Köhler and Dinnerstein did not use a pre-inspection, control measure of the apparent width of the standard stimulus as a baseline for evaluating the size of the after-effect.[1] They regarded this to be the difference between the observer's (*O*'s) point of subjective equality (*PSE*) and the objective width of the standard after exposure to an inspection-stimulus either wider or narrower than the standard. Wertheimer later pointed out that control measurements preceding the inspection-period should be taken to avoid inclusion of constant errors from pre-inspection judgments.[2] At present, most investigators follow this suggestion and use pre-inspection, control judgments in studies of kinesthetic after-effects.

An assumption implicit in the use of pre-inspection, control judgments is that these judgments themselves do not influence the after-effects. It is the purpose of this paper to investigate the validity of this assumption.

The possibility that the pre-inspection judgments do influence the after-effects needs to be considered, since in making the pre-inspection judgments a set may be established to do the same thing when *O* is later exposed to the same situation, despite the intervening inspection-stimulus. The question is whether anything happens during the pre-inspection judgments to influence the effect of the inspection-stimulus on the post-inspection judgments.

Method. Two groups of 39 undergraduates, each composed of 28 men and 11 women, were used as *O*s. The *O*s of Group I made eight pre-inspection judgments of the width of the standard stimulus; those of Group II made no pre-inspection judgments.

The apparatus consisted of three wooden blocks: (1) a standard with parallel sides, 1½ in. wide, 60 in. long, and 2 in. thick; (2) a wedge-shaped block, 60 in. long, and 2 in. thick, which varied in width from 1 in. to 2⅞ in. at the rate of 1/32 in. per inch; and (3) an inspection-stimulus, with parallel sides, ½ in. wide, 24 in. long, and 2 in. thick. The three blocks were mounted on tables 36 in. high and

---

* Received for publication October 28, 1960. This research was supported by a grant from the National Science Foundation.

[1] Wolfgang Köhler and Dorothy Dinnerstein, Figural after-effects in kinesthesis, in *Miscellanea Psychologica Albert Michotte*, 1947, 196-220.

[2] Michael Wertheimer, Constant errors in the measurement of figural after-effects, this JOURNAL, 67, 1954, 543-546.

parallel to each other. The standard- and inspection-blocks were at the side of *O*'s non-preferred hand and the variable at the side of *O*'s preferred hand. The variable block was fitted with an H-shaped rider within which *O* placed his thumb and forefinger while making equality-judgments.

Pre-inspection (Group I) and post-inspection judgments were made by finding a width on the wedge-shaped block which felt equal to the width of the standard block held in the other hand. Half the judgments were made from a point ¼ in. wider than the standard and half from a point ¼ in. narrower than the standard; the order of judgments alternated in a descending-ascending order. Just prior to making four post-inspection judgments, the *O*s in both groups were given a 90-sec. inspection-period during which they moved the thumb and forefinger of the non-preferred hand back and forth along the inspection-stimulus. The *O*s wore blindfolds during the judgment- and the inspection-periods.

Results. Group-comparisons were based on the post-inspection trials. There was no significant difference between the groups in variability on post-inspection trials. For purposes of analysis, the mean of the first descending and the first ascending judgment was considered Trial 1, and the mean of the second pair of judgments was called Trial 2. The results of the experiment are presented in Table I.

TABLE I

DEVIATION OF POST-INSPECTION JUDGMENTS FROM OBJECTIVE MATCH

*O*s	Trial 1	Trial 2	*M*
Group I	.15 in.	.11 in.	.13 in.
Group II	.19 in.	.18 in.	.185 in.
M	.17 in.	.145 in	

The mean of Group II is larger, though not significantly so when tested by an analysis of variance (F = 1.58, 1 and 76 *df*). The direction of the difference tends to support the hypothesis that pre-inspection judgments make *O* resistant to the effect of the inspection-stimulus. The difference between groups is somewhat greater on Trial 2, but the trial by group interaction does not reach significance (F = 2.51, 1 and 76 *df*). The difference between trials is significant at the 5% level (F = 4.72, 1 and 76 *df*), indicating a decay in the size of the after-effect between trials.

Summary. In a study of the effect of pre-inspection, control judgments on kinesthetic after-effects, the post-inspection judgments of two groups were compared. One group made 8 judgments of the apparent width of a block 1½ in. wide before being exposed to an inspection-block 0.5 in. wide. A second group made no pre-inspection judgments. No statistically significant difference between the groups was found on the post-inspection judgments. The direction of the observed difference suggested larger kinesthetic after-effects for the group without the pre-inspection judgments.

APPARATUS

A SIMPLE AND RELIABLE APPARATUS FOR THE MEASUREMENT OF PAIN

By ERNEST G. POSER, McGill University

The apparatus to be described provides a useful alternative to thermal and electrical devices currently employed in studying the threshold and tolerance-levels of pain.[1] It consists of a standard clinical sphygmomanometer with a pressure gauge calibrated to 300 mm./Hg. The pain-stimulus (Fig. 1) is delivered by a series of 94 pointed projections 7 mm. in height and arising from a flat acrylic base roughly 4½ ×2¾ in. in size. This stimulator consisting of four sections is sewn into the sleeve of the sphygmometric cuff. The reason for the stimulator's being in sections rather than in one piece is to allow for greater flexibility. When the cuff is in position the projections come to rest against the medial surface of *S*'s upper arm. The projections are sharp enough to discourage most *S*s from tolerating more than 250 mm./Hg. of pressure but not so sharp as to cause skin-laceration at 300 mm./Hg. beyond which point the instrument should not be used.

A controlled amount of pressure can now be exerted upon the cuff by inflating it with the standard air-pressure bulb of the sphygmomanometer. This procedure however, introduces uncontrolled variables arising from fluctuations in speed and grip-pressure used by different examiners when they inflate the cuff. For this reason we have introduced a source of compressed air in lieu of the pressure-bulb (see Fig. 2).

The air tank has a capacity of 2000 lb. of pressure and is fitted with a one-stage reduction-valve. The airflow is regulated by a pressure-compensated flowmeter calibrated to output-levels ranging from 0 to 15 liters per min. In this way a constant air-supply is delivered to the cuff through an 11½ ft. latex rubber tube with a 6-mm. lumen. We have found one liter per min. to be the optimal air-supply for our purposes. At this setting the cuff is inflated at the rate of 12.5 mm./Hg. per sec. Air-input is by way of a metal T-junction in the rubber tube as shown in Fig. 2. To inflate the cuff the experimenter simply occludes the outlet from the T-junction with his thumb thus forcing air into the cuff. Lifting the thumb will instantaneously release pressure from the cuff.

---

[1] The work here described was partially supported by Dominion Provincial Mental Health Grant No. 604-5-57 and the Verdun Protestant Hospital Research Fund.

To determine threshold and tolerance-levels of pain, *S* is instructed to report when pressure changes to pain. A reading is taken, whereupon the cuff is immediately deflated. Then it is again inflated with *S* being instructed to report the point at which pain becomes intolerable. About 200 *S*s can be investigated in this manner without refilling the tank.

The relationship between this mechanical method and those using thermal or electrical means of producing cutaneous pain has been investigated by Clark and Bindra,[2] who found significant intercorrelations among the three methods, somewhat higher at the tolerance-level than at the thresholds. These investigators reported immediate retest reliabilities for the pressure technique ranging between 0.86 and 0.91. We have administered the method to various groups of university students obtaining rank-order retest correlations between 0.75 and 0.85 with a mean time-interval of 12 days between test and retest. The method has been found useful in drug studies, psychodiagnostic classification, and the socio-psychological study of differences among ethnic groups.

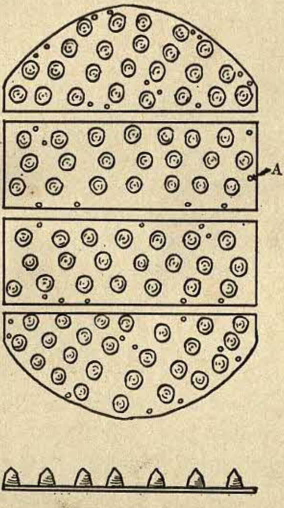


FIG. 1. STIMULATOR FOR PAIN
A = holes throught which the stimulator is sewed to the cuff.

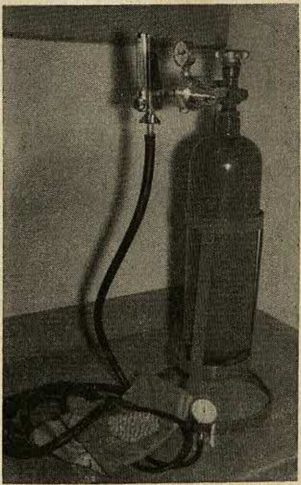


FIG. 2. APPARATUS FOR PAIN
Pressure tank, stimulator, and accessories.

---

[2] J. W. Clark and Dalbir Bindra, Individual differences in pain thresholds, *Can. J. Psychol.*, 10, 1956, 69-70.

APPARATUS NOTES

A SIMPLIFIED CONSTANT-CURRENT SHOCKING CIRCUIT

Constant-current shocking systems usually have involved electronic circuitry or the use of high resistances in series with *S*. The present system provides the high degree of current-regulation yielded by electronic methods with the economy and convenience of the series-resistance method, from which it differs only in the substitution of a capacitor for the series-resistor (Fig. 1a).[1]

The principle involved is that the total impedance of a resistor (*S*'s body) and a capacitor in series do not add like two resistances in series, but instead like two vectors at right angles (Fig. 1b). That is, the total impedance, in ohms, equals $(R_S^2 + R_C^2)^{\frac{1}{2}}$, where R_S is the resistance of *S* and R_C is the impedance, in ohms, of the capacitor. Thus, if the capacitor has 10 times the impedance of *S*, then reducing the *S*'s resistance to zero would increase the current by only about 1%. The system requires alternating current. For 60 cycles, the *rms* current in µamp. is approximately 0.000377 *EC*, where *E* is the shock voltage and *C* is the capacitance in micro-micro-farads (*MMF*). For example, if $E = 665$ v., the current will be ¼ µamp. per *MMF*.

The basic circuit, shown in Fig. 1a, requires only a high-voltage transformer and a capacitor. The capacitor may be a small ceramic-disk type, rated at several thousand volts (*e.g.* Centralab Series DD60, which is rated at 6000 v.). If desired, two or more of these can be wired in parallel to give a specific current needed. In practice, the basic circuit is usually modified (a) by adding fuses and a microammeter in series with the secondary of the transformer, and (b) by using several fixed capacitors (each with a switch) plus a variable capacitor, allowing the current to be varied continuously over a wide range.

A typical variable-current circuit, with a nominal range of 20-1000 *MMF*, or approximately 5-250 µamp. is shown in Fig. 2. The variable capacitor should be a transmitting type, such as Cardwell type MO-165-BE (Burstein-Applebee Co., Kansas City, Missouri). The fuses are for the protection of *E* and should be of the instrument-type, such as Littlefuse, series 361000, 1/500 amp. The alternating-current microammeter, essential for purposes of calibration, is manufactured and sold by Simpson Electric Co., Chicago, Ill. (Model 49, range 0-200 microamperes). When adjusting the current, the electrodes or grids should be short-circuited.

A feature of the system which may be desirable is that current is not maintained when the *S*'s resistance becomes very high, which happens when *S* steps on or off a grid-bar. If the current remained nearly constant, the current *density* (*i.e.* current per unit area of skin surface) would become extremely high, thus punishing locomotion. The present system should minimize punishment for locomotion while providing good current-regulation when *S*'s feet are not moving. Experimental comparisons between the methods, however, have not been made.

---

[1] Constant current is obviously inappropriate where there is appreciable current leakage in the grid or other shocking apparatus. Before employing constant current, therefore, it is well to check the apparatus with an ohmmeter; resistance should be many times higher when it is empty than when it contains an animal.

If the transformer, capacitors, and meter are located at some distance from the animal, an appreciable current loss may occur in the cable, causing the current actually delivered to be less than that shown on the meter, as well as impairing the current-regulation. Most of the loss can be avoided by using two separate wires, rather than a cable, to conduct the current to the animal. Alternatively, the entire shocking apparatus may be placed near the animal, with the current controlled by varying or switching the voltage applied to the primary of the transformer, or by means of relay-contacts in series with the *S*.

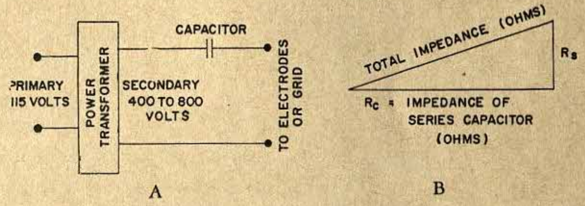


FIG. 1. DIAGRAMS (A) OF THE BASIC CIRCUIT, AND (B) OF THE IMPEDANCE-RELATIONSHIPS

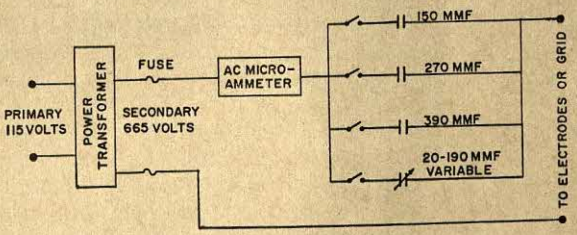


FIG. 2. A PRACTICAL CIRCUIT

University of Mississippi WILLIAM F. CROWDER

AN IMPROVED MECHANICAL FISH-LEVER

Recent interest in the behavior of submammalian organisms has presented a number of technical problems, one of which has been the design of a lever to be used in the operant conditioning of aquatic forms. Several such devices have been described,[1] including one which makes use of an electronic circuit.[2] The lever described in this paper operates more reliably than the previous mechanical devices, and is very sensitive. It also is easy and inexpensive to build. This lever has been employed in extended experiments on two types of fish, ranging in weight from less than 1 gm. to more than 100 gms.[3] It operates consistently and is very sturdy.

---

[1] J. V. Haralson and M. E. Bitterman, A lever-depression apparatus for the study of learning in fish, this JOURNAL, 63, 1950, 250-256; W. A. Deterline, An automatic device for the investigation of operant behavior in the fish, this JOURNAL, 69, 1956, 291-295.

[2] Nicholas Longo and M. E. Bitterman, Improved apparatus for the study of learning in fish, this JOURNAL, 72, 1959, 616-620.

The moving parts of the lever consist of the armature assembly of a general telephone-type relay. The pivot in these assemblies has been engineered commercially to provide reliable and extensive service. The armature-yoke (A) is extended with a piece of plastic. The tip of the armature-lever makes contact (B) with a brass screw mounted in the plastic. A piece of silver is soldered both to the tip of the armature-lever and to the brass screw to insure good contact. One end of a stainless-steel shaft is soldered to the free-swinging armature (C). An appropriate target is attached to the other end of the shaft. The lever is normally closed; that pressure against the target breaks the contact between the armature-lever and brass screw. The sensitivity of the lever can be adjusted by a counterweight (D) attached to the armature, and a piece of sponge-rubber (E) pressed against the armature.

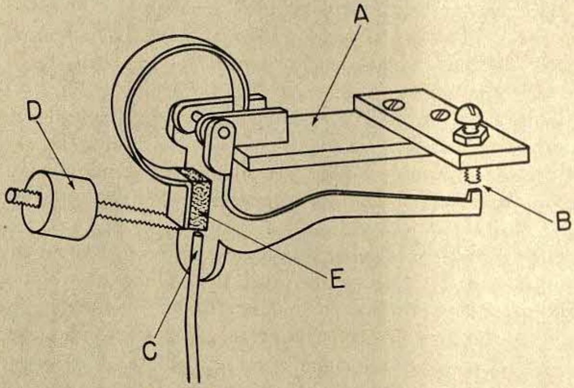


FIG. 1. DIAGRAM OF THE LEVER
(A, armature-yoke; B, contacts; C, armature; D, counterweight; E, foam-rubber.)

Harvard University

JERRY A. HOGAN
PAUL ROZIN

---

[3] J. A. Hogan, Motivational aspects of instinctive behavior in *Betta splendens*, Unpublished Doctoral dissertation, Harvard University, 1961; Paul Rozin, Some aspects of regulatory behavior in the goldfish, *Carassius auratus*, Unpublished Doctoral dissertation, Harvard University, 1961.

AN AUTOMATIC APPARATUS FOR THE CONTINUOUS METHOD OF LIMITS

To obtain threshold measurements under conditions in which a minimal amount of time is a consideration—thresholds must be determined quickly—the automatic apparatus described here was devised. The procedure used is the continuous method of limits. A variable stimulus is set to an intensity below threshold and is increased in intensity. *S* signals when he first perceives a difference. The stimulus-intensity continues to change, however, in the same direction for a brief interval, which is varied in duration from trial to trial in a pseudo-random manner. Then the direction of the change is reversed and *S* again signals when he no longer perceives a difference. After another brief, pseudo-random interval, the direction of change is again reversed, and so on for as many series and reversals as desired.

This method differs from the standard method of limits in that it permits the convergence of the stimulus-series about the threshold. It differs from Békésy's "method of adjustment" in that it assures independence of successive determinations by interposing pseudo-random amounts of stimulus-change between successive series.[1] The use of a converging series, in addition to permitting the rapid determination of the threshold, minimizes adaptation by keeping the stimulus close to the threshold.

The apparatus consists of a motor driving a cam in which three grooves are cut, with different amounts of space between each two. A microswitch is activated each time a groove passes its position. It is thus activated in repeating cycles of three variable intervals. When *S* responds by pressing a button, indicating that he perceives or no longer perceives the stimulus or a difference, relays are so placed that the stimulus-intensity is reversed in direction of change the next time a groove hits the microswitch. As few as three variable delays thus interact with *S*'s variability to produce a set of delays which appears random over long sequences of responses.

As presently used, this device controls the motor on a variac, which regulates the voltage delivered to a projector-bulb. The bulb projects a spot onto an illuminated background. When *S* responds, indicating that he first sees the spot, the voltage continues to increase at a rate fixed for the particular motor-driven variac (4-v. per sec.), until the next time a groove passes the microswitch. The variac-motor is then reversed and the intensity of the stimulus decreases. The grooves on the cam are separated by spaces in the ratio of 3:4:5, and the cam is rotated at a speed of 10 rpm.

This apparatus has been used on the centrifuge of the Navy Aviation Medical Acceleration Laboratory and reliable brightness-difference thresholds were determined in periods of 90 sec. or less, while *S* was exposed to various levels of acceleration.

Cornell Aeronautical Laboratory, Inc. MYRON L. BRAUNSTEIN

---

* This method and apparatus were developed for research sponsored by the Office of Naval Research under Contract NOnr-3457(00) with the Cornell Aeronautical Laboratory, Inc.

[1] George von Békésy, A new audiometer, *Acta Oto Laryngologica,* 34, 1947, 411-422.

AN APPARATUS FOR THE STUDY OF BINOCULAR RIVALRY

The device described here, and pictured in Fig. 1, was designed in an attempt to increase binocular rivalry.[1] The two targets, consisting of identical Plateau spirals, are rotated in opposite directions by a variable-speed motor, through a simple gear train, and are viewed through the eye-pieces of a Brewster stereoscope. It was anticipated that the contradictory rotary movements, plus the conflicting explosive-implosive effects, would decrease fusion and increase rivalry. The research possibilities of the apparatus have not as yet been explored. The apparatus has, however served admirably for intriguing and effective demonstrations. These moving targets appear to be more forceful in producing rivalry than do the usual static ones.

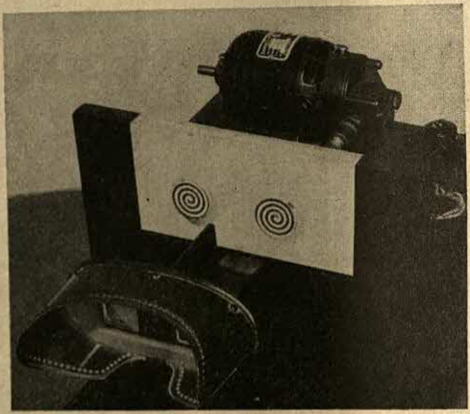


FIG. 1. APPARATUS FOR BINOCULAR RIVALRY

The Pennsylvania State University WILLIAM M. LEPLEY

---

[1] This device was built by John L. Kobrick, a graduate assistant in the Department.

NOTES AND DISCUSSIONS

MEASURES OF INFORMATION BASED ON OVERLAPPING COUNTS FOR STUDYING SEQUENTIAL DEPENDENCIES

In 1954, I published a note which pointed out the existence of certain equations relating cell-frequencies based on overlapping counts of sequential data.[1] Such overlapping counts are used in studying sequential dependencies in sequences of responses (or of stimulus-response combinations); for a full discussion of the problem of sequential dependencies and the use of the information statistic in studying it, see Shannon, Frick and Miller, Newman, and my note in 1954.[2] The remainder of this note assumes general familiarity with the underlying problem and technique.

The only use for those equations which was apparent in 1954 was to reduce the total number of sequences which must be tabulated. At that time, the Miller-Madow correction for bias in estimates of measures of amount of information and test of significance for those estimates had not yet been published; both the correction for bias and the test of significance depend on the number of degrees of freedom (*df*) in the data-matrix.[3] But one implication of the equations to which my note in 1954 called attention is that the *df* of data-matrices based on overlapping counts are far fewer than conventional formulas for calculating *df* would indicate. This means that the Miller-Madow correction applied to measures of information calculated from overlapping counts will seriously overcorrect unless

---

[*] This work was conducted by Project MICHIGAN under Department of the Army Contract DA-36-039-SC-78801, administered by the U. S. Army Signal Corps. I am very grateful to Dr. W. M. Kincaid, who solved the gap-problem, and to Dr. G. A. Miller, who spotted a major error.

[1] Ward Edwards, Methods for computing uncertainties, this JOURNAL, 67, 1954, 164-170. An example of an overlapping count: suppose a subject produces the sequence of responses *ABAABABAAB*. This sequence contains 3 *ABA*s, 2 *AAB*s, and 1 *BAB*. Addition of the last two numbers indicates that the sequence *AB* occurred three times (note that the first occurrence of *AB* is lost in thus obtaining counts of shorter sequences by adding counts of longer ones; in practical applications, this loss is negligible).

[2] C. E. Shannon, A mathematical theory of communication, *Bell Sys. tech. J.*, 27, 1948, 379-423, 623-656; F. C. Frick and G. A. Miller, Statistical behavioristics and sequences of responses, *Psychol. Rev.*, 56, 1949, 311-324; E. B. Newman, Computational methods useful in analyzing series of binary data, this JOURNAL, 64, 1951, 252-262.

[3] G. A. Miller and W. G. Madow, On the maximum likelihood estimate of the Shannon-Wiener measure of information, Air Force Cambridge Research Center: Technical Report 54-75, August 1954.

formulas for calculating *df* are used which take these equations into account. Since this point has not previously been discussed, this note discusses it and presents the appropriate formulas.

Degrees of freedom for uncertainties. An example will introduce the problem and indicate the nature of its solution. Consider a sequence of events of length n, in which any of b different possible events may appear (b is very much smaller than n). We are interested in overlapping counts of all sequences of length s, $s - 1$, . . . , 1. Of course the counts at shorter sequence-lengths are obtained from those of longer sequence-length by addition. Thus if $b = 2$ and $s = 4$, we might have counted from the data the number of occurrences of the sequences *ABAA, ABAB, AABA, BABA,* among others. We now wish to find out how many times the sequences *ABA* occurred. We can do so in either of two ways; by adding $N_{ABAA} + N_{ABAB}$, or by adding $N_{AABA} + N_{BABA}$. Furthermore, it should not matter which of these two ways we choose; the results should be the same (except for negligible effects associated with sequences which appear at the beginning or the end of the total sequence). Consequently, we can write:

$$N_{ABAA} + N_{ABAB} = N_{AABA} + N_{BABA} (= N_{ABA}). \qquad [1]$$

Furthermore, there will be one such equation for each sequence of length $s - 1$. It looks as though there are half as many equations (for $b = 2$) as unknowns, so that only half the sequences need be counted. This is true, but only if the sum-check on the over-all count is used as one of the equations.

Now, what about *df?* Continuing with the example $b = 2$, $s = 4$, calculation of the sample-uncertainty in sequences of length s (or shorter lengths) is perfectly straightforward; but that sample-uncertainty will be a biased estimate of population-uncertainty. If we want to correct it using the Miller-Madow correction, we can easily do so if we know the proper *df.* Examination of the data-matrix shows us b^s (in this example 16) cells in the part of the matrix on which that uncertainty is based. But it would be incorrect to use $b^s - 1$ as the appropriate *df.* The problem can be solved by reverting to the definition of *df.* The number of *df* is simply the number of cells which can be filled in freely; in this case, 8. Of course, 8 is also the appropriate number of *df* for any test of the significance of the difference between the uncertainty calculated from these data for sequences of length 4 and the maximal uncertainty possible for such sequences, whether a chi-square calculated directly from the data or the Miller-Madow test is used.

The general problem and its solution. The preceding example illustrates the nature of the problem and of its solution, but does little justice to the degree of complexity that the problem may attain. The sequence of events being considered may include two or more different kinds of events. For example, responses and consequences or outcomes produced by them may alternate, as in an experiment on probability-learning. The number of different possible events of any single kind need not be two; nor need that number be the same for different kinds of events. Finally, the sequence whose uncertainty is to be measured may not include all events intervening between the first and the last event considered. For example, an experimenter might be interested in an uncertainty calculated from the present response, the immediately previous response and its outcome, and the outcomes but not the responses of the last three trials before that. It is as easy, though not ordinarily as appropriate, to calculate uncertainty from a sequence which includes gaps as from a continuous sequence. Equation [2] below will give the correct *df* for any calculation of uncertainty based on overlapping counts except for certain kinds of sequences with gaps; they are discussed in the next section.

A formal general definition of the problem will introduce notation as well as leading to the solution. Define a set M of classes of events; that set has m members. Class of events J contains j different members, class of events K contains k different members, and so on. The data are a sequence of n cycles of events; each cycle will be called a trial. Each trial contains exactly one event from class J, one event from class K, one event from class L, and so on, in that order.

To perform the overlapping count appropriate to calculation of an uncertainty, a set Q of structures which might appear in the sequence of data must be defined. The uncertainty to be calculated is the average uncertainty with which, if we entered the sequence of data at a randomly chosen trial, we could predict which member the set Q would begin (or end, or any consistent rule) at that trial. The set Q is a sequence of classes of events, short relative to n. In this section, only sequences without gaps will be considered; the following section examines sequences with gaps. That means that all classes of events intervening between the first and the last class included in Q must, in this section, also be included in Q. For example, the set M might contain two classes, responses (R) and the outcomes (O), occurring in that order. A set Q might then have the form *RORОR*, but forms such as *OOR* are excluded from this section. Note that any particular class of events may be included in Q more than once, if Q extends (as it almost always does) over more than one trial. Call the number of occurrences of classes of events included in Q s; in the *ROROR* example, $s = 5$.

Once the set Q has been chosen, its members must be listed. In the example, if each response may be A or B and each outcome may be a or b, the members of Q include *AaAaA, AaAaB, AaAbA,* and so on; in this example Q has $2^5 = 32$ members. Next, the number of cycles over which Q extends must be determined; call that number c. In the example, $c = 3$. Of course, c must be very small relative of n for any calculation of uncertainties to be meaningful. Next, Trials 1, 2, . . . , c of the sequence of data are examined, and a tally is entered opposite the member of Q of which those trials in the sequence are an instance. This process is repeated for sequence of Trials 2, 3, . . . , $c + 1$, again for 3, 4, . . . , $c + 2$, and so on until all trials in the sequence of data have been examined. This process will generate a table of data having $n - c + 1$ entries. An uncertainty can be calculated from that table in the usual way.[4] Incidentally, in general such calculations should be made from values of $N \log_2 N$ rather than $p \log_2 p$. The procedures are logically equivalent, but the procedure based on *N*s is easier and avoids serious rounding-off errors likely to occur if *p*s are used.

---

[4] For details, see Fred Attneave, *Applications of Information Theory to Psychology,* 1959, 46-57.

Once an uncertainty has been calculated, it is usually appropriate to apply the Miller-Madow bias correction to it and sometimes appropriate thereafter to test the hypothesis that the obtained corrected uncertainty is significantly different from the maximal possible uncertainty for that table of data, using either the Miller-Madow test or a chi-square test. To do any of these things, it is necessary to know the appropriate *df*. By definition of *df* that *df* will always be the total number of cells in the table minus the number of *independent* equations relating cells in the table to one another. In the class of sequences *Q* suppose that elements of class *J* appear x times, elements of class *K* appear y times, and so on; $x + y + \ldots = s$. Then the number of cells in the matrix of data will be $j^x k^y \ldots$ The number of independent equations can be found as follows. First, define a set *T* of sequences of events which might be found in the data. *T* is defined by the fact that it is the longest set of sequences of events whose frequencies in the data can be determined from the matrix of data defined by *Q* in two or more different ways. For example, if *Q* is *ROROR, T* is *ROR*. If *Q* is *OROR, T* is OR. If *Q* is *RRRR* (no outcomes considered), *T* is *RRR*. For sequences without gaps (and for sequences with experior gaps only), each class of events appears exactly once less often in *T* than in *Q*. The point of set *T* is that for each of its members one equation like Equation [1] can be written. So the number of such equations will be equal to the number of elements in *T*. If all but one of these equations is given, the remaining one can always be deduced. Other ways of generating equations from *Q* can be found, but such equations always (for sequences without gaps) turn out to be deducible from those generated by set *T*, and so are not independent. The sum-check, however, is independent of all these equations. Consequently the total number of independent equations is also equal to the number of elements in *T*. The number of elements in *T* is, of course, $j^{x-1}k^{y-1} \ldots$ So, since *df* is the number of cells minus the number of independent equations,

$$df = j^x k^y \cdots - j^{x-1}k^{y-1} \cdots = j^{x-1}k^{y-1} \cdots (jk \cdots - 1). \qquad [2]$$

For the specific case, often encountered in psychology, in which each class of event has the same number b of members (*i.e.* $j = k = \ldots = b$), Equation [2] reduces to

$$df = b^s - b^{s-m} = b^{s-m}(b^m - 1). \qquad [3]$$

Equations for still more specific cases can be found by substituting appropriate values for, for example, b and m into Equation [3]. In particular, *df* for sequences of two classes of binary events are $(3/4)(2^s)$, and for a single class of binary events are $1/2(2^s)$.

Sequences with gaps. Sometimes a set of sequences *Q* which includes gaps might be of interest. For example, you might believe that *S* uses the last three outcomes but only the immediately previous response in deciding what response to make, and so you might be interested in the uncertainty of a set *Q* of the form *OOROR*. To highlight the existence of a gap, a better notation might be *O·OROR·*, where each dot indicates that an element of a trial is being ignored. (Note that the preceding discussion tacitly ignored the outcome following the present response whenever the response-outcome example was used.)

It is appropriate to divide gaps into two classes: interior gaps and exterior gaps. An exterior gap is one in which, if a particular class of events is ignored in a particular trial included in Q, it is also ignored either in all trials included in Q preceding or all following that trial. An interior gap is any gap which is not an exterior gap. Thus the sequences $O \cdot OROR \cdot$, $O \cdot OR \cdot R$, and even $O \cdot O \cdot O \cdot \ldots R \cdot R \cdot R$ have only exterior gaps. A short sequence including an interior gap is $RO \cdot OR$; if only responses are considered, a shorter one is $R \cdot R$.

Equations [2] and [3] apply to all sequences which include only exterior gaps. They do *not* apply to sequences with interior gaps, except that they provide a lower limit for *df* for such sequences. I have been unable to develop an analytic equation for the number of *df* for sequences which have interior gaps, but Wilfred Kincaid (personal communication) has developed an algorithm which will give the right answer. The algorithm which follows applies to sequences which have only one kind of event in them ($m = 1$). Kincaid has also developed an algorithm suitable for sequences of any number of different kinds of events with interior gaps, but it is tedious, complicated, extremely difficult to explain, and, in view of the low probability that anyone might want to calculate the uncertainty of such sequences, not worth publishing.

Here is Kincaid's algorithm for sequences of only one kind of event which have interior gaps:

(1) Write down the sequence. Example: $RR \cdot R \cdot R$

(2) Replace filled positions with numerals. Example: 12.4.6

(3) Find all subsets not including the numeral 1 that cannot be transposed (rewritten) to begin with 1 without using numerals not used in Step 2. Example: $2 \cdot 4$, $4 \cdot 6$, $2 \ldots 6$, $2 \cdot 4 \cdot 6$ cannot be so transposed. If the example had been $RRRR \cdot R$ at Step 1, only $234 \cdot 6$, $23 \ldots 6$, and $2 \ldots 6$ cannot be transposed; 23, $2 \cdot 4$, 234, $34 \cdot 6$ $3 \ldots 6$, and $4 \cdot 6$ can all be transposed. All single-numbered subsets can always be transposed.

(4) Tabulate the length of each non-transposable subset found in Step 3. For the $12 \cdot 4 \cdot 6$ example, there are three two-element subsets and one three-element subset. For the example of $1234 \cdot 6$, there are one two-, one three-, and one four-element subset.

(5) Define u_i as the ith subset length found in Step 4, and v_i as the number of different non-transposable subsets having that length.

(6) Using notation defined earlier in this paper along with that defined in Step 5:

$$df = b^{s-1}(b - 1) + \sum_i v_i(b - 1)u_i. \qquad [4]$$

Thus, if the number of possible responses at any trial is $3 (b = 3)$, then $RR \cdot R \cdot R$ has $3^3(3 - 1) + 3(3 - 1)^2 + 1(3 - 1)^3 = 38$ *df*. Similarly, the sequence $RRRR \cdot R$ has 190 *df*.

Note that sequences in which interior gaps are regularly spaced (example: $R \cdot R \cdot R$) have no non-transposable subsets not including the numeral 1, and so Equations [2] and [3] apply to such sequences directly and without modification.

This is also true if whole trials are regularly gapped in sequences which have more than one kind of event, but is not true if there are regular gaps for one kind of event but no corresponding ones for the other kinds of events.

This whole question of gaps, and particularly of interior gaps, is rather academic. Sequences including gaps are never appropriate for the calculation of components of transmitted information, and it is very difficult to think up instances in which simple uncertainties for sequences which include interior gaps would be of interest.[5]

Estimates of transmitted information. So far the discussion has focussed on uncertainties, which are calculated from a single matrix of data. An apparently more complicated problem arises in correcting the bias of estimates of transmitted information, which require several data-matrices based on the same counts for their calculation. But this apparently more difficult problem is really no problem at all. Any amount of transmitted information may be written as the algebraic sum of two or more uncertainties, and the appropriate *df* are simply the negative of the algebraic sum of the appropriate *df* for the component uncertainties, calculated from Equation 2 or 3.[6] Whenever overlapping counts and sequential dependencies are at issue, McGill's caution that events intervening between the event considered as the source of information and the event considered as the receiver must be held constant is of crucial importance; estimates of transmitted information which do not conform to that caution are usually uninterpretable.[7] That is why sequences which include gaps are not appropriate for use in calculating estimates of transmitted information.

University of Michigan WARD EDWARDS

INACCURACIES IN THE SEMANTIC COUNT OF LORGE AND THORNDIKE

Thorndike and Lorge's *The Teacher's Word Book of 30,000 Words,* is often cited for its counts of relative frequency of usage of English words.[1] Recently we had occasion to use one of these counts, the semantic of 'S' count. This count also appears in part in three other publications that will be mentioned below. Comparing data from the four publications, we found many discrepancies and errors. We wish to call the existence of these inaccuracies to the attention of users and prospective users of these lists.

---

[5] W. J. McGill, Isomorphism in statistical analysis, in H. Quastler (ed.), *Information Theory in Psychology,* 1955, 56-62.

[6] For justification see Attneave, *op. cit.,* 63-67.

[7] McGill, *op. cit.,* 57.

[1] E. L. Thorndike and Irving Lorge, *The Teacher's Word Book of 30,000 Words,* 1944.

The initial semantic count, prepared in the 1930s, was based on an original sample of over two million running words taken from a variety of texts. For some words a supplementary sample of the same size was also used. For each word except some of the most common, every occurrence was studied to determine its meaning in context, and the meaning was then classified according to the list of meanings given in *The Oxford English Dictionary*.[2] The findings were published in 1938 as a work of six volumes, *A Semantic Count of English Words*.[3] This publication gives the relative frequency of usage of each meaning of each word listed. Elaboration of the semantic count was continued, and the 'S' count in *The Teacher's Word Book of 30,000 Words,* published in 1944, was based on fuller information than was the original report. The 'S' count, like the other counts in the same volume, gives only the total frequency of usage for each word, rather than breaking down the usage of each word among its various meanings. In 1949 Lorge published a supplement to the original report, giving the semantic information about the 570 commonest words in the language.[4] Finally, in 1953, West published *A General Service List of English Words,* which is essentially an abridgment and simplification of the semantic count of Lorge and Thorndike.[5]

The size of the sample employed for the count is given variously in the four publications. In the preface of *A Semantic Count of English Words,* Thorndike wrote that most of the frequencies were based on the original sample of "about two and a quarter million words."[6] In *The Teacher's Word Book of 30,000 Words* we read that the sample for most words "included about four and a half million words but for some it included only two and a quarter, in which case the number of occurrences is doubled before entry in the 'S' column."[7] In *The Semantic Count of the 570 Commonest English Words*, Lorge noted that the frequencies for some words were based on the original sample, while for others they were based on both the original and the second sample. Here, however, he stated that the "two presumably equivalent samples [consisted] each of about two and a half million words."[8] "Whenever the relative frequencies are based on less than the total sample of five million words, estimates of the frequency per five million were made."[9] These estimates are found in a separate table, pp. x-xiii. Finally, *A General Service List of English Words* gives the frequencies per five million words.[10] There is thus disagreement as to whether the semantic count was based on four and a half million or five million words.

---

[2] *The Oxford English Dictionary,* Vols. 1-13, 1933.

[3] Lorge and Thorndike, *A Semantic Count of English Words,* Vols. 1-6, 1938.

[4] Lorge, *The Semantic Count of the 570 Commonest English Words,* 1949.

[5] Michael West, *A General Service List of English Words,* 1953.

[6] Lorge and Thorndike, *op. cit.,* 1938.

[7] Thorndike and Lorge, *op. cit.,* 1944, 253.

[8] Lorge, *op. cit.,* v.

[9] *Ibid.,* vi.

Beyond this fundamental ambiguity, there remain further discrepancies among the four publications. Table I presents for a few words the data on frequency of usage, taken from each of the four sources in which the word is listed. Two of the sources, that of Lorge and Thorndike and that of Lorge, indicate for each word the size of the sample employed, and this is shown in the table. The first word in the table, 'bird,' was not counted in the original sample, so all data concerning it presumably come from the second sample, yet the three sources give quite different figures for this word. The next three words, 'color,' 'dead,' and 'high,' were also counted only in the second sample. For these three words, West reproduces the figures of Lorge, but Thorndike and Lorge give discrepant figures for two of the words and omit the word 'dead.' It should be noted that the value given for a word by Thorndike and Lorge is sometimes greater than that given by the other sources, as in the case of the word 'high,' but it is sometimes less, as for the word 'color.' The last five words in the table were counted in both the original and the second sample. For such words, West usually doubled the values found in Lorge and Thorndike, but in some cases arithmetic errors apparently occurred (*e.g.* 'put' and 'state'). West does not give any reason for using the original sample in these cases instead of using the total sample, which was available to him. Note also that

TABLE I

FREQUENCIES OF OCCURRENCE OF WORDS ACCORDING TO FOUR PUBLICATIONS BASED ON THE SEMANTIC COUNT

	Lorge-Thornkike (1938)	Thorndike-Lorge (1944)	Lorge (1949)	West (1953)
Words	2,225,000	4,500,000	5,000,000	5,000,000
bird	——	348	168†	840
color	——	958	1086†	1086
dead	——	——	613†	613
high	——	3110	2809†	2809
piece	329*	755	737‡	658
put	851*	2082	2082‡	1602
state	3174†	6221	11211‡	7348
table	359*	677	677‡	718
write	774†	861	2167‡	1548

* Counted in the original sample. † Counted in the second sample. ‡ Counted in both samples.

---

[10] West, *op. cit.*, vii.

for three of the last five words in the table, the figures given by Thorndike and Lorge do not agree with those given by Lorge. The value of 11,211 given by Lorge for 'state' appears improbably high, not only in relation to the other figures given in the table for 'state' but also in relation to the value of 1564 given in the Lorge Magazine Count.[11]

To get a general indication of the comparability of the frequencies given in West and in Thorndike and Lorge (these being the two most widely available publications based on the semantic count), we took the word at the top of every tenth page in West and compared its frequencies in the two sources. It should be noted that most of the words included in West are of rather high frequency of occurrence, and their frequencies should therefore be rather stable for comparable samples of textual material. Of the 58 words selected for the comparison, four do not appear in the 'S' columns of Thorndike and Lorge. Of the remaining 54 words, seven have identical frequencies in the two books. To obtain a common measure of the magnitude of differences of frequencies in the two sources, the difference for each word was expressed as a percentage of its frequency in West. For all 54 words, the median difference was about 9%. For six cases in our sample, the two books give frequencies that differ by more than 50%. Thus, while this comparison indicates that the two publications are in general agreement concerning the frequencies of words in the semantic count, the proportion of sizable discrepancies is not negligible.

University of California, Berkeley

MARK R. ROSENZWEIG
DAVID MCNEILL

PSYCHOPHYSICAL MEASUREMENT OF THE SPIRAL AFTER-EFFECT

The spiral after-effect has recently attracted attention because of the findings by Price and Deabler that patients suffering brain-damage fail to report it.[1] Other authors have questioned the validity of Price and Deabler's original findings.[2] All investigators of this phenomenon have used subjective reports of the presence or absence of the after-effect. A method for making accurate measurements of the rate of change of perceived size is highly desirable. We have developed an apparatus which makes such measurement possible and are now in the process of exploring the diagnostic usefulness of this new technique.[3]

---

[11] Thorndike and Lorge, *op. cit.*, 262.

[1] A. C. Price and H. L. Deabler, Diagnosis of organicity by means of the spiral after effect, *J. consult. Psychol.*, 19, 1955, 299-302.

[2] H. C. Holland and H. R. Beech, The spiral after effect as a test of brain damage, *J. Ment. Sci.*, 104, 1958, 466-471; Perry London and J. H. Bryan, The influence of instructions on spiral after effects, *Amer. Psychologist*, 13, 1958, 335 (Abstract).

The spiral after-effect is elicited by having a subject (*S*) fixate a rotating disk on which has been drawn an Archimedes spiral with its center at the point of rotation. After the rotating disk has been fixated for a preparatory interval (usually 20–30 sec.), the disk is stopped suddenly and *S* is asked to report what he sees. When a clockwise spiral is rotated clockwise, it appears to shrink; when it is abruptly stopped, the spiral appears to expand and to come closer to *S*. If this spiral is rotated counter-clockwise, it appears to expand during rotation and to shrink and to recede from *S* when rotation is stopped.

Our method for measuring the rate of change of perceived size involves alternate presentation of the rotating spiral and a changing stimulus in a mirror-tachistoscope. The changing stimulus is produced electronically in the form of a circular, oscilloscopic figure which expands or contracts at accurately controlled rates. The signal-generator which produces the circular figure consists of a sine-wave generator, a phase-shift network, and a two-channel modulator unit having a variable amplification-factor which is controlled by a variable-rate, timing circuit. The signal-generator is calibrated in percentage-units based upon an initial size of circle of 3 in. The settings used range from a contraction of 20% per sec. through an expansion of 20% per sec. in steps of 2% per sec. There are thus 21 different settings, including a setting for no change in size of circle. *S* views the rotating spiral for 10 sec., and the circle for 2 sec. These two stimuli are alternated tachistoscopically, each alternating cycle

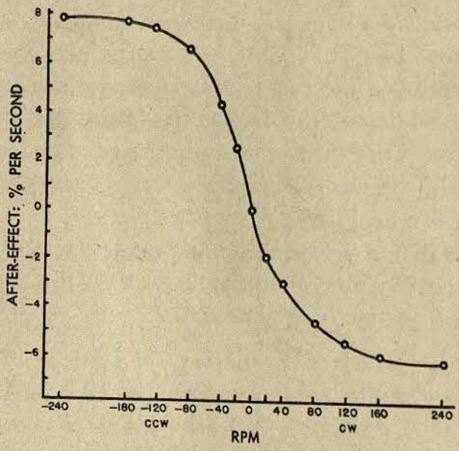


FIG. 1. SPIRAL AFTER EFFECT AS A FUNCTION OF ROTATION
These data were obtained with a 4-turn, 8-in. spiral viewed at a distance of 5 ft.

---

[3] Financial assistance for instrumenting this research was provided in part by the South Carolina Academy of Science.

comprising a single trial with no rest-period between successive trials. *S* reports expansion or contraction of the circle each time it appears. The method of limits can thus be used to determine that rate of expansion or contraction of the circle which produces the subjective experience of size-constancy. By this method a highly reliable mean for the after-effect can be obtained in between 50 and 60 trials. This apparatus has been named the *metamegethograph* (change + size + graph), abbreviated *MMG*. A name for the after-effect measured in this way is needed to discriminate clearly between it and that obtained by the earlier methods.

Fig. 1 shows the results obtained with four normal *S*s on an 8-in. diameter, 4-turn, counter-clockwise outward spiral viewed at a distance of 5 ft. Six different clockwise and six counter-clockwise speeds were used. The standard error of the points on the curve is between 1% per sec. near the center of Fig. 1 and 2% per sec. near the ends. The after-effect appears to increase with negative acceleration as a function of speed of rotation. The maximal negative after-effect (apparent shrinkage) is between 5 and 6% per sec., while the maximal positive after-effect is between 7 and 8% per sec. These values are probably near the maximal rates.

VA Hospital THOMAS R. SCOTT
Columbia, South Carolina RUFUS E. MEDLIN

PHENOMENAL CAUSALITY IN CONDITIONS OF INTRA-SENSORY AND INTERSENSORY STIMULATION

Michotte observed, in 1946, that if a visual stimulus is brought toward another which moves away from the first upon apparent impact, the situation is described as being one in which the second stimulus is in some way set into motion by the first.[1] This phenomenon he has called 'causality' and it is observed when *S*s' verbal descriptions of the stimulus-situation imply a causal relationship between the two successive events. He interprets the phenomenon as being a basic element of perception since the effect has been observed with very large numbers of naïve *S*s and it persists even when they are made fully aware of the nature of the experimental conditions.

In previous experiments into problems of perception in conditions of 'intersensory' stimulation, it has been suggested that information about the positions of stimuli in space may be handled by a single integrating framework located in the central nervous system.[2] If this is the case, it should follow that the 'causality' experience may not be limited to the visual modality but available in other senses which handle spatial information. Further, it should be evident in 'intersensory' conditions also when, for example, the first stimulus is a light, the second a sound.

---

[1] A. E. Michotte, *La perception de la causalité*, 1946, 1-295; 2nd. ed., 1954, 1-306, esp. 41-413.

[2] G. H. Fisher, (1960a) Intersensory localisation in three modalities. *Bull. Brit. Psychol. Soc.*, 41, 1960, 24A f.; Intersensory elements of phenomenal space, *Proc. XVIth Int. Cong. Psychol.*, 23, 1960. 19.

Accordingly experiments were conducted with 51 naïve *S*s in both 'intrasensory' (first and second stimulus in same modality and 'intersensory' (first and second stimulus in different modalities) conditions. The apparatus used for these experiments was rather different from that used by Michotte and has been described fully in a separate paper.[3] In these experiments the presentation of an 'intrasensory' or 'intersensory' stimulus-pair was randomized among the *S*s. One hundred presentations of every stimulus-pair, movement being randomized between right to left and left to right, in each of the three 'intrasensory' and six 'intersensory' conditions, were given irrespective of what kind of response the *S*s made, all of these responses were noted fully. Briefly, the results were as follows: (a) 5 *S*s *never* made a verbal response which could be interpreted as being causal; and (b) 46 make a causal response during the course of the experiments.

Of Group b 4 *S*s made a causal response *on the first trial,* the remainder, at some stage during the course of the experiment, did so. Once a causal response had been made, however, *no S* ever changed back to making a non-causal response of any kind, but simply either continued to repeat his original response or on a few occasions changed to a different causal response (*e.g.* 'push' to 'kick,' often in an exasperated way) and although the experiment was completed with all *S*s, there often seemed to be little point in so doing. In other words, although often many trials are required before the 'causal' response is made, once it is made it is not changed.

This finding appears to be further support for the presence of a central integrating system of the kind previously hypothesized, it also appears to be a further illustration of the way in which the categories of language structure the categories of perception. What appears to be happening here is that an attempt is being made to find suitable words with which to describe the stimulus-events. In this case, words in the 'causal' category appear to be most suitable and once a response in this category is made it is adhered to subsequently *even when the kinds of spatial stimuli pre-*

---

[3] Fisher, Some problems associated with the design and construction of apparatus suitable for presenting spatial cues simultaneously in different modalities, unpublished MS.; Cf. also Thomas Natsoulas, Principles of momentum and kinetic energy in the perception of causality, this JOURNAL, 74, 1961, 394-402.

sented are changed. Unfortunately, to our knowledge, no similar work has been conducted on very young pre-verbal children with whom a different kind of response would need to be observed, but we are presently designing experiments in which we hope to be able to observe sensorimotor responses in stimulus-situations of this kind.

University of Durham GERALD H. FISHER
England

NEW ILLUSIONS OF PITCH

An immense literature exists on geometric-optical illusions, but auditory illusions, particularly those involving pitch, are as yet totally undescribed. I exclude here, of course, auditory hallucinations and the pathological cases of "false hearing." I have in mind particularized, normal illusions of musical pitch. I shall describe how I discovered two cases of this kind.

I have a masonite disk of 16-in. diameter and ⅛-in. thickness. On it I bored three concentric circles of holes equally distributed. The inner circle has 32 holes; the middle, 48 holes; and the outer, 55 holes. Mathematically the ratio 32:55 represents an unmusical interval approximately in the middle between the musical intervals of the Major Sixth (3:5) and the Natural Seventh (4:7). To have the experiences which I shall describe, one must, however, approach in an unsophisticated way, that is, either keeping uninformed of the mathematical facts or keeping them out of mind.

Experiment I. Blow a blast of air alternately, in the slow tempo of, say, a religious hymn, at Circles 32 and 55. Instruct the listener to determine the musical interval. A violinist thinking professionally would at once say "That is not a musical interval." To my nonchalant mind, however, a song of long ago, beginning with a Major Sixth, enters my consciousness and I cannot help telling myself that it is a Major Sixth. Various persons invited and listening under similar conditions have given the same report, "It is a Major Sixth."

Experiment II. Mount the disk on a motor which spins it from 0 to 20 rotations per sec. Choose a different motor-speed to forget the absolute pitch of Experiment I. Now blow into the three circles (32, 48, and 55) in the same tempo as previously used as many times as the listener requests. When I listen, the idea at once comes to mind that the three tones I hear, if heard simultaneously, would constitute the so-called Dominant-Seventh Chord (except that the Major Third is absent). I hear a Tonic plus its Fifth plus its Natural Seventh; *i.e.* the pitch which in the first experiment was a Major Sixth is now a Seventh.

Explanation. The brain tends to convey to consciousness adjusted and musical intervals which it substitutes for unmusical ones. That of course is *the fundamental neurological* fact, the source of the grand illusion without which we could never have a truly esthetic experience when we listen to a piano or to an orchestra that gives us *the score mistuned* in a tempered tuning which the musical world has accepted. The explanation of the illusions reported above is as follows.

(I) The brain, with esthetical efficiency, makes a substitution for a ratio of such high terms as 32 and 55, which is 55:32 = 1.719. It has to choose between neighbors, *i.e.* either 5:3 instead of 55:32, or 7:4 instead of 55:32. Since the numbers 3 and 5 are smaller than 4 and 7, the brain selects the former two numbers as ratio-terms. We compute 5:3 = 1.667 as a measure versus 1.719 for the size of the illusion. We thus are conscious of the Major Sixth instead of an unmusical interval that is physically, by a considerable amount, larger than the Sixth.

(II) The brain rejects the multiple ratio 32:48:55. It has to choose between 6:9:10 and 4:6:7. The sum of the former is 25, and the sum of the latter three is 17. The brain selects the latter as the smaller terms. Thus the listener hears the interval 32:55 as if it were 4:7, the Natural Seventh. We compute 7:4 = 1.750 versus 1.719 as the measure of the illusion in this case. Other persons under similar conditions, I find, have the same illusion that I have. Let me state it in the following words: The addition of the Fifth to the other two pitches (32 and 55) produces the illusion of the tone 55 being something different from what it physically is. It is the addition of the Fifth that forces the brain to make a choice different from the choice which the brain made in the first experiment.

5308 Danville St.
Springfield, Virginia
MAX F. MEYER

CRITICAL FLICKER FREQUENCY IN RELATION TO CATTELL'S CULTURAL FAIR INTELLIGENCE SCORE

In 1956, Landis and Hamwi reviewed and reanalyzed the available data on the relation between the critical flicker frequency (*CFF*) and intelligence.[1] They concluded that the only statistically reliable correlations reported up to that time—by Tanner and by Colgan[2]—could be treated as random fluctuations. In none of the studies reviewed was a measure of intelligence used that could be considered culturally free. It would seem logical, however, that the nearer one gets to achieving a cultural free measure of intelligence the more the total variance of the test will reflect basic neuro-physiological processes involving sensory imput-systems. This assumption led us to examine the relation between *CFF* and scores on the Cattell Culture Fair Intelligence Scale (*CFIS*), a test developed to minimize the influence of cultural experience.

---

[1] Carney Landis and Violet Hamwi, Critical flicker frequency, age, and intelligence, this JOURNAL, 69, 1956, 459-461.

[2] W. P. Tanner, A preliminary investigation of the relationship between visual fusion of intermittent light and intelligence, *Science*, 112, 1950, 201-203; C. N. Colgan, Critical flicker frequency, age, and intelligence, this JOURNAL, 67, 1954, 711-713.

In one study, the *CFIS* (Scale 2, Form A) was administered to 98 fifth-grade children who already had taken the Wechsler Intelligence Scale for Children (*WISC*). On the basis of *CFIS* scores, 42 children (14 each with high, middle, and low scores) were selected for the *CFF*-test, which was measured with a Grass Photo Stimulator equipped with a special shield. Stimuli were presented by the method of limits. The *WISC* total *IQ*s were not significantly correlated with *CFF* ($r = -0.09$). The *CFIS* was significantly correlated with *CFF* ($r = 0.43$; $p < 0.01$), supporting the hypothesis under test.

In a second study, the Otis Quick Scoring Mental Ability Test (Gamma, Form F_m) and the *CFIS* (Scale 3, Form A) were administered to 70 college sophomores. The *CFF* was determined as before, and, in addition, the *S*s were given an auditory discriminatory task (involving 36 paired tones ranging from 1000–8500$\sim$). *CFIS*-scores were significantly correlated with *CFF* ($r = 0.243$; $p < 0.05$), but not with the auditory-discriminative scores ($r = 0.225$), although the correlation approached significance. The Otis scores were significantly, but *inversely*, correlated with CFF ($r = -0.329$; $p < 0.01$) and with the auditory-discriminative scores ($r = -0.540$; $p < 0.01$). The auditory scores and *CFF* were not significantly correlated ($r = -0.205$). Again, the results relating *CFF* to the *CFIS* support the hypothesis under study. The significant inverse relationships of the Otis scores with *CFF* and with the auditory-discriminative scores are difficult to explain.

Texas Christian University

Ernest S. Barratt
Marjorie Clark
James Lipton

A PHYSICAL MODEL OF PERCEPTION AS AN INTEGRATING PROCESS

For our purpose, we may define 'information' as the body of signals that must necessarily be transmitted and received to complete a certain process. The question is then raised whether the receiver is able, to a certain extent, to integrate in time. If, on repetition, the content of information as detected by the receiver does not vary, we may suppose that none of its significant signals is less intense than the background noise present in any type of transmission. In the opposite case, however, if any signal less intense than the background noise is cut out, information *will* be affected. In fact, any one of the components of the background noise will, if integrated over a sufficient length of time, be equal to zero, whereas any signal less intense than the background noise, if integrated over a sufficiently long period, will give a detectable result, greater than zero. This phenomenon, as well as the possibility of detecting signals below the noise-level, are well known in electronics. The interesting fact is that this suggests a model for biological perception. The problem becomes extremely complex since it cannot be taken for granted that all subliminal signals are essential to information, but it would be very important to be able to state that even some of these signals are essential. This would appear to be perfectly acceptable, since any stimulus acts on living cells in a different manner according to whether it is preceded by another stimulus. This is, in fact, a form of integration.

Department of Physics
University of Buenos Aires
Argentina

ANDREA LEVIALDI

THE 1962 MEETING OF THE NATIONAL ACADEMY OF SCIENCES

The National Academy of Sciences held its ninety-ninth annual meeting at the Academy building in Washington on April 23-26, 1962.

Thirteen of the twenty-three members of the Section of Psychology were present: F. A. Beach, E. G. Boring, Leonard Carmichael, Robert Galambos, J. P. Guilford, E. R. Hilgard, D. B. Lindsley, W. R. Miles, N. E. Miller, Carl Pfaffmann, C. P. Richter, L. A. Riggs, S. S. Stevens. Hilgard was elected chairman of the Section to replace Lindsley whose term expires.

Dr. Frederick Seitz, physicist of the University of Illinois, was elected President to succeed Detlev W. Bronk, who is retiring after nine years of service.

The Academy elected George A. Miller of Harvard University to membership. During the year the Section of Psychology had lost Arnold Gesell by death.

The Daniel Giraud Elliot Medal for the most meritorious work in zoology or paleontology published each year was presented to Donald R. Griffin of Harvard University for his work on echolocation of bats.

The following papers of psychological import were read:

D. R. Griffin, A. D. Grinnell, J. J. G. McCue, D. Cahlander, and F. A. Webster: "Neural discriminations achieved by bats in echolocation."

J. P. Guilford, Philip R. Merrifield, and Arthur Gershon: "Some primary abilities in the areas of nonverbal divergent production."

Neal E. Miller, Kay S. Gottesman, and John E. Holmes: "Effects of chemostimulation of brain and of bacterial endotoxins on eating and drinking."

S. S. Stevens: "Sensation measurement by inverse judgments."

Roger J. Williams, Richard B. Pelton, and Frank L. Siegel: "Individuality as exhibited by inbred animals; Its implications for human behavior."

D. W. Woolley: "Alterations of learning ability caused by changes in cerebral serotonin and catechol amines."

In a brief program the Academy received from the Equitable Life Assurance Society the gift of the new north wing of the Academy building as a Hall of the Life Sciences. The wing is not quite finished and will be dedicated in 1963 at the centennial meeting of the Academy.

Harvard University EDWIN G. BORING

THE 1962 MEETING OF THE AMERICAN PHILOSOPHICAL SOCIETY

The Annual General Meeting of the American Philosophical Society was held in Philosophical Hall in Independence Square, Philadelphia, April 26-28, 1962.

The Society was addressed by Herbert A. Simon of the Carnegie Institute of Technology on "The architecture of complexity: Some common properties of complex systems," by John C. Lily of the Communication Research Institute at Miami on "Vocal behavior of the bottlenose dolphin," and by John T. Emlen, Jr., of the University of Wisconsin on "The display of the gorilla." These were the papers of greatest import to psychologists.

The fourth Karl Spencer Lashley Award "in recognition of useful and significant work in the field of neurobiology" was made to Philip Bard of The Johns Hopkins University. The presentation was made by Leonard Carmichael. Previous awardees have been Rafael Lorente de Nó, Heinrich Klüver, and Lord Adrian.

Robert R. Sears, Professor of Psychology at Stanford University, was elected to membership. The psychologists present at the meeting were Frank A. Beach, Edwin G. Boring, Leonard Carmichael, Wolfgang Köhler, Walter R. Miles, and Curt P. Richter.

Harvard University EDWIN G. BORING

FIFTY-EIGHTH ANNUAL MEETING OF THE SOCIETY OF EXPERIMENTAL PSYCHOLOGISTS

The fifty-eighth annual meeting of the Society of Experimental Psychologists was held at the University of Michigan on April 20 and 21, 1962. Arthur W. Melton, Chairman of the Society for the year, served as Chairman of the meeting.

Members present were: Bartley, Beach, Bray, Brogden, Deese, Egan, Fitts, Galambos, Garner, Geldard, E. J. Gibson, J. J. Gibson, Harlow, Helson, Humphreys, Hurvich, Kappauf, Kendler, Kennedy, Melton, Meyer, Morgan, Neff, Newman, Shepard, Spence, Underwood, Volkmann, and Wickens.

H. W. Hake, I. J. Hirsh, G. A. Kimble and N. H. Mackworth were elected to membership in the Society, bringing the total membership to 64 members and 20 fellows.

Members presented reports of research in progress at scientific sessions held the afternoon of April 20 and the morning of April 21.

The Warren Medal for 1962 was awarded to James Olds "for his development and effective exploitation of intracranial electrical stimulation as a fruitful method in the analysis of positive reinforcement."

The Society accepted the invitation of Brown University to meet in Providence in 1963. Harold Schlosberg was elected Chairman for 1962-63.

Princeton University JOHN L. KENNEDY

SEVENTH CONGRESS OF THE INTERAMERICAN SOCIETY OF PSYCHOLOGY

The Seventh Congress of the Interamerican Society of Psychology was held in Mexico City from December 19 to 23, 1961, under the sponsorship of the Social Security Institute of Mexico, National University of Mexico, Rama Mexicana de la SIP, and the Centro Investigaciones Socials. Approximately 650 members and visitors from 11 countries (Argentina, Brazil, Chile, Colombia, Cuba, Guatemala, Mexico, Panama, Peru. and the United States) attended the Congress.

All the sessions of the five-day meeting were held at the Centro Medico of the Seguro Social, recently completed. Papers and addresses were translated into English, Portuguese, and Spanish over earphones as they were given, hence they could be understood by all. The Organizing Committee was under the able direction of Dr. Guillermo Davilla G. of Mexico City.

Ninety-five papers were presented under the following sections: "Culture and personality," "Experimental psychology," "Applied psychology (educational and industrial)," and "Psychology and mental health." Five plenary meetings were devoted to addresses by Dr. Gustave Gilbert (U.S.) on "The needs for a comprehensive bio-social theory of personality"; by Dr. Erich Fromm (Mexico), on "The revolutionary character structure"; by Dr. Carlos Alberto Séguin (Peru), on "Language and communication"; by Dr. Abraham Maslow (U.S.), on "The scientific study of values"; and by Dr. Robert B. Malmo (Canada), on "Contributions of experimental psychology to the clinical interview."

The program also included two symposia on "The development of the personnel subsystem in industrial and military man-machine systems," with Dr. Jerome Ely (U.S.) as coördinator; and "Transcultural impact at the U.S.-Mexican border," with Dr. Rogelio Díaz-Guerrero as coördinator.

Our Mexican hosts provided an excellent program of social activities, including a gala night of entertainment by the National Dancers of the Seguro Social.

Arrangements have been made to publish, in English and Spanish, the proceedings of the Seventh Congress under the leadership of Dr. Guillermo Davila G.

The officers of the Society elected for 1962-1963 were: Dr. Jose Bustamante, President; Dr. Gustave M. Gilbert, Past-President; Dr. Harold H. Anderson, President-Elect; Dr. Victor D. Sanua, Executive Secretary for North America; Dr. Rogelio Díaz-Guerrero, Executive Secretary for Mexico and Central America; Dr. Fernanda Monasterio, Executive Secretary for South America; and Dr. Rafael Núñez, Treasurer. Additional members of the Society's governing board are the following Vice Presidents: Dr. Wayne H. Holtzman, for North America, and Dr. Carlos Alberto Séguin, for South America.

The Eighth Congress of the Interamerican Society of Psychology is scheduled for April 2 to 6, 1963, and will be held at Mar Plata, Argentina, under the sponsorship of La Plata University.

Yeshiva University VICTOR D. SANUA

BOOK REVIEWS

Edited by T. A. RYAN, Cornell University

Learning Theory and Behavior. By O. HOBART MOWRER. New York, John Wiley & Sons, Inc. 1960. Pp. xii, 555. $6.95.

A new spirit of inquiry fast is beginning to characterize contemporary thought in learning. The fervor of the great controversies of the thirties and forties has been subsiding gradually as experimental evidence relating to the issues has accumulated; and the raw enthusiasm which led to strong allegiance with one or another of the major systems has given way to a more calm appraisal of various points of view. As a result, old ties are being severed; there is a growing willingness to look anew to the vast literature and to make fresh starts in search of better conceptions.

This transition, to a large extent, has taken the form of a rebellion against the Hullian tradition. Hull, himself, of course, led the way. His 1952 *Behavior System* contrasted sharply in some fundamental respects with his 1943 *Principles of Behavior.* Recently, Spence, for many years the strongest spokesman for the Hullian position, also has introduced some major modifications in his views. And now Mowrer, who had defected earlier from Hull, presents the third major work in the past decade to reflect this trend.

Learning Theory and Behavior, the first book of a two-volume set, is concerned with basic processes. The book itself is divided into two parts. The first seven chapters present, in historical perspective, a detailed and highly personalized account of the development of Mowrer's new theoretical position. The remaining chapters are devoted to a comparison of this position with other contemporary views and to the application of the theory to a variety of basic phenomena.

Mowrer's early defection from Hull stemmed primarily from his attempt to provide a unified account of the facts of punishment and avoidance. Hull had attempted to reduce all learning to the formation and strengthening of stimulus-response connections through the agency of need reduction, making classical conditioning a special case of the operation of the law of effect. The resistance of avoidance and punishment phenomena to analysis in terms of this monistic conception led Mowrer, in 1947, to the formulation of two-factor theory: autonomic responses (emotions) are learned in Pavlovian fashion and skeletal responses (solutions) in Thorndikian fashion. Avoidance-learning is a two-stage process, the conditioning of *fear* as a function of *CS–US* pairing, and the strengthening of an instrumental response as a function of fear-reduction. Punishment also involves the conditioning of fear; proprioceptive feedback from the punished response comes to arouse fear, and cessation of response provides the fear-reduction necessary for the establishment of the (negative) habit.

Mowrer's revision of two-factor theory stems primarily from a reanalysis of the action of reward in terms of the principles already developed to account for punishment. In laying the groundwork for the new theory, the author draws a sharp distinction between two types of secondary reinforcement: Type-1 (relief) occurs with the termination of a fear-producing stimulus, and Type-2 (hope), with the onset of a stimulus which has been associated with primary drive-reduction (reward). At the root of the new version of "habit" is "hope."

In the "revised two-factor theory" instrumental learning disappears entirely, and all learning is reduced to Pavlovian conditioning of emotional reactions either to exteroceptive or to proprioceptive stimuli. Drive *change* is asserted to be an essential condition for learning; drive-increment as well as drive-decrement may reinforce (in the Pavlovian manner). Stimuli which are contiguous with drive-increments acquire secondary-drive (fear-arousing) properties, while stimuli which are contiguous with drive-decrements acquire secondary-reinforcing (hope-arousing) properties. Instrumental behavior is a by-product of the conditioning of proprioceptive stimuli. Reward, like punishment, serves not to alter specific instrumental (*S–R*) tendencies, but to condition emotional reactions to the feedback from the instrumental response. Positive feedback leads to the perpetuation of the response, while negative feedback leads to its cessation.

Mowrer believes that this position takes a middle ground between stimulus-response and cognitive theory, representing a "creative synthesis thereof" which is "sounder than either of these more extreme positions." This conviction is difficult for a critical reader to share. Although the traditional *S–R* view of habit has been abandoned, *reinforced response* remains an essential condition for change in instrumental behavior. The response whose probability of occurrence is to be altered must occur if its feedback is to occur, and it must be contiguous with drive-decrement or drive-increment if its feedback is to acquire facilitative or inhibitory properties. The theory thus faces all of the difficulties which confront the traditional *S–R* view. Furthermore, the difficulties are compounded by the problem of response *selection*. Once initiated, a response is facilitated or inhibited by its sensory consequences, but how is the response initiated to begin with? The author prefers to deal with the problem in *Learning Theory And the Symbolic Processes,* the companion volume to the present work. The solution, however, already has been offered in a previously published paper (*Psychol. Rev.*, 63, 1956, 114–128). It is based on the principle of scanning. In a state of need, the organism scans its behavioral repertory (by making covert or overt responses) until it hits upon a response the sensory feedback of which has been associated in that situation with drive-decrement. This account is extremely difficult to reconcile with data on latency. Latencies of a small fraction of a second simply do not yield to such analysis. Furthermore there seems to be no principle by which account may be taken of progressive decline in latency in the course of learning.

The very core of the theory, independently of the general failure to transcend the limitations of the traditional *S–R* view, also is open to question. The basic assumption that stimuli which are contiguous with drive-change inevitably acquire drive or reinforcing properties leads to the inflexible position of leaving no room for an analysis based upon the important discriminative (cognitive) functions which such stimuli may serve. Consider, for example, the situation in which a warning signal is introduced in the course of a Sidman-avoidance schedule. Although the stimulus is paired repeatedly with shock, animals do not avoid it (do not "fear" it). On the contrary, they utilize it, postponing response until it appears. Similarly, as the experiment of Bitterman, Feddersen and Tyler demonstrates, the discriminative properties of stimuli paired with reward may far outweigh any reinforcing value they may acquire.

These experiments are cited in the book. Indeed, Mowrer's presentation is characterized by an earnest effort to treat a wide range of data and a willingness to deal with evidence which challenges the theory. His treatment of these data, however, is not very incisive; and it often is marred by cumbersome excursions which confuse the issue. The book is too long, the discourse too tedious. A vast literature is reviewed (often mis-viewed); there are extensive quotations from books, journals, correspondence, classroom meetings, and conversations; footnotes abound. The reader swims in a sea of redundancy, sections that should have been paragraphs, paragraphs that should have been sentences.

In conclusion, it should be noted that this work represents a prodigious effort. It is difficult to imagine another single source which brings together in detailed fashion such an impressive array of primary literature in the field. The theory, however, does not come up to the effort.

Bryn Mawr College R. C. GONZALEZ

Learning Theory and the Symbolic Processes. By O. HOBART MOWRER. New York, John Wiley & Sons, 1960. Pp. xi, 473. $8.50

Although published as a self-contained monograph under its own title, this is actually the continuation of a work whose first part came out a few months earlier under the title *Learning Theory and Behavior.* A summary of the points made in the companion volume forms the first chapter of *Learning Theory and the Symbolic Processes,* but anybody who reads the second volume without having read the first will be, if anything, rather worse off than somebody who begins an average serial story with the second installment.

In the first volume, Mowrer expounded a new theory of behavior which he called "revised two-factor theory." Its crux was the assertion that all learning depends on the acquisition, through conditioning by contiguity, of punishing (drive-inducing) or rewarding (drive-reducing) properties by response-produced feedback stimuli. In the second volume, Mowrer goes on to extend this theory to such topics as latent learning and latent extinction, imitation, thinking, probability and statistics, social learning. One of his chapters—probably the best of them—consists of an adaptation of his article "The psychologist looks at language." (*Amer. Psychologist,* 9, 1954, 660-692.)

It must be said at the outset that this is an interesting and important book. If it were written by almost anybody else, one might well leave it at that, but since its author is Mowrer, from whose previous work so many of us have derived great intellectual and even esthetic pleasure, one must also register some disappointment.

The disappointment concerns both the organization and the content, which must inevitably be interconnected. Mowrer has evidently been wrestling with these problems for many years, and he makes it clear that he is still not satisfied with the formulations he has here written down. One can, in fact, picture the publisher snatching the manuscript from the author's unwilling hands and running off with it to forestall more chopping and changing and rethinking. The text is riddled with expressions of the author's still unsettled misgivings about what he has so far put to paper, of afterthoughts, of footnotes referring to experimental data and ideas that became available to the author at the last moment, and even of frank explanations that certain passages were inserted when the book was in galley proof.

References to galley proofs and such prosaic features of book production have a rather displeasing effect; it is rather as if an operatic tenor broke off in the middle of an impassioned aria to tell the audience what happened while he was gluing on his whiskers.

The result is that the book forms a rather disconnected medley. It does not consist of a sustained argument, but rather of a series of essays of varying style and quality, some of which record the author's earnest but inconclusive thinking out loud.

As for the content, there is one lacuna that is especially regrettable. When Mowrer first put forth the ideas underlying his revised two-factor theory in 1956, they were recognized by many readers as significant and challenging. The stress on proprioceptive feedback was supported by some experimental data, especially from Eastern Europe, and dovetailed neatly with cybernetics; but there was one troublesome defect. The theory might explain why an animal goes on performing a response immediately after that response has produced a reward. But it did not seem to explain why the time taken for the first performance of the response should diminish when the animal is placed in the experimental situation on successive occasions.

About half way through *Learning Theory and Behavior,* Mowrer shows himself to be fully aware of this problem. He promises that a solution to the problem will be offered in the forthcoming volume (the one under review), and he asks the reader meanwhile to "take it 'on faith' that a solution to this basic problem *is* available." The reviewer could not recollect encountering any solution to the problem after he had read through the second book, and, on going back over the relevant chapters, he found nothing more than brief indications of two or three alternative lines along which a solution might be found. This is a great pity, since it is difficult to evaluate the revised two-factor theory, let alone its extensions to the symbolic processes, unless one can see in some detail how this crucial difficulty might be overcome.

In general, Mowrer is here exemplifying a most interesting trend in contemporary behavior theory, namely a willingness to face squarely the questions raised by the unique flexibility and rationality of the more complex forms of human behavior, while resisting any suggestion that these properties make something like vitalism or mentalism necessary. A parallel trend is apparent, significantly enough, in contemporary Soviet psychology.

Mowrer comes out, for example, with the assertion that the image must find a place in objective behavior theory and that attempts to do without such a concept will not work. It is noteworthy that Beritov, probably the most microphysiological of Pavlov's living successors, has likewise felt compelled by some of his recent findings to introduce the notion of a "conditioned image." Several Western authors, such as Skinner, Hebb, J. S. Brown, and J. G. Taylor, have been doing something similar, although they have not all brought themselves to use such terminology. If, however, "images" are introduced as intervening variables, the shady previous history of the image concept makes it especially imperative that its operational definition and its relations with observable antecedent and consequent variables should be carefully worked out and specified. But Mowrer does not do this. Nor does he present in closely argued detail his precise reasons for insisting that the "image" must be reinstated.

When one looks at the texture of his theory more closely, several other perplexities emerge. For example, he explains latent learning and latent extinction by appealing to mediating stimulus-producing responses. He explains how, if conditioned affective properties of mediating stimuli should become altered, overt behavior might change on a future occasion. If one is to follow out the implications of revised two-factor theory, it is easy to see why mediating responses representing desirable events should be performed, since the feedback stimuli consequent on them will have acquired positive affective tone. But it is not so clear why mediating responses representing dangerous or noxious states of affairs should occur at all to guide behavior, since their feedback stimuli will be negatively toned.

The first volume of the pair was, among its other merits, particularly valuable as a source of bibliography. Mowrer gave a good coverage of experimental work relating to the points he was discussing in it. Most of this work is in any case American, but he also took care to mention enlightening East European studies, particularly those of Konorski's school. The present volume is much less successful as a bibliographical sourcebook and review of the literature. One special obstacle to complete coverage is the fact that a high proportion of recent experimentation on the symbolic processes has been carried out in the Soviet Union and in Geneva, Switzerland, but this work is not reported, nor is due account taken of it.

These limitations—and it would be pointless to pretend that they are negligible ones—stand out because of the high standards that Mowrer sets himself, and most of them arise directly out of the virtues for which he is so widely esteemed: his willingness to pick out and grapple with the loose ends in his own thinking, his eagerness to acknowledge and make use of ideas suggested by others, his readiness to accept the challenges that experimental data from the most diverse sources offer to his own theory.

So none of the reasons for disappointment must be construed as an excuse for not reading the book with the utmost respect. The science of psychology will always be deeply in Mowrer's debt for his key role in building up the learning-theory movement and in showing learning theory to be not a narrow field of specialization but something with an intimate relevance to all areas of psychology. Whenever Mowrer writes on learning theory, every psychologist must consider carefully what he says or be deemed remiss in his duty of keeping abreast of contemporary psychology.

Boston University

DANIEL E. BERLYNE

The Psychology of Character Development. By R. F. PECK with R. J. HAVIGHURST, and RUTH COOPER, J. LILIENTHAL, and D. MORE. New York, John Wiley & Sons, 1960. Pp. xx, 267. $6.50.

This is an important book. In tough-minded contrast to oceans of a priori, armchair literature on morality, it reports a relatively successful, ambitious attempt at a longitudinal, quantified, empirical study in depth of a sample, in a small Midwestern town, of 34 children, born in 1933, with respect to their personality dynamics, social development, and maturity of character from age 10 to 17 yr. as affected by their social environment.

The 10-page Preface briefly reviews the history of scientific work in this area which concluded that character was largely a function of situational circumstances, hence not very predictable, and little influenced by 'character-building' agencies such as church, school, and Scouts, thus contradicting 'popular opinion' about the generality of moral character, its origins, and growth. The authors of the present study note in passing that "in method as well as in subject matter this has been an exploratory research" whose findings, because of the small number of cases, must be regarded as provisional. Ten chapters and a 58-page Appendix constitute the substantive report. Technical details, the clanking of the research machinery, are largely relegated to the Appendix.

The research, planned and executed by a team trained in psychology, sociology, and education, as reported in this book "represents the end product of years of exploratory study by a good many people." The theoretical framework is somewhat eclectic and considerably psychoanalytically oriented. The basic data were extensive and intensive case studies made with an enormous number and variety of observational and measurement techniques ranging from physical anthropometrics to projective, sociometric, interview, and rating techniques.

Five character types—none pure in any individual—evolved from the research. They constitute the major conceptual, higher-level abstractions from the theoretical frame of reference and enormous masses of data. These in successive stages of growth are as follows: (1) The *Amoral*—often called clinically the "psychopathic individual," follows his whims and impulses without regard to how it affects others. (2) The *Expedient*—primarily self-centered, considers other people's welfare and reactions only to gain his own ends. (3) The *Conforming*—has one general internalized principle: to do what others do and what they say one 'should' do. (4) The *Irrational-Conscientious*—behaves according to rigidly internalized principles of 'right' and 'wrong,' a 'blind' superego at work. (5) The *Rational-Altruistic*—the highest level of moral maturity, has a strong, firm superego but tests, modifies, and applies its directives in accordance with social reality and ultimate consequences for other people.

In the words of the Preface: "If this book appears to present a simple theory, in reasonably clear terms, with some provisional evidence to support it, this is a consummation for which we are profoundly grateful." In general, its conclusions are in many ways diametrically opposed to those of the "Character Education Inquiry" which were that character is stable and hence predictable at least from age 10 to 17 yr. for this population-sample and presumably throughout life.

The research team faced enormous difficulties, but with a high degree of sophistication. The hazards of being culture-bound, the conceptual and operational definitions of many variables observed and inferred from behavior, the possibilities of circular reasoning, the problems of obtaining quantitatively comparable data from the children under study, from their parents and other adults in the community, their peers, their teachers and school records, the possibilities of Baconian Idols of the Den (built-in halo effects), the quantitative handling of individual personality and character profiles, problems of reliability and validity of measurement—all of these were realistically and painstakingly faced and solved with some, and somewhat varying, degrees of success.

That moral character is formed early and chiefly by parental behavior is not a new idea, since it is supported by a sizeable body of other previous research. That only about a quarter of the population-sample, at a rather generous estimate, reached the Rational-Altruistic level of behavior while 75% remain in varying degrees moral cripples is a new and depressing finding in its implications for the mental health of individuals and society. The 'character-building' agencies in general and at best have reinforcing rather than originating roles in character development.

To raise questions that are negatively critical of such a complex study is not difficult. A minor irritation, incidentally, is the consistent misspelling of Riesman's name. The laboratory-oriented experimental psychologist will not be happy with the necessarily limited controls possible in such a study. To interpret a demonstrated curvilinear relationship between two variables in terms of a Pearson r is not 'kosher,' nor is interpreting the difference between an r of 0.64 and 0.58 as psychologically real; the logic of inferring causality from certain correlations—these are, when all is said and done, at worst only minor slips in interpreting data generally painstakingly collected, analyzed, and cross-checked. Solving the problem of quantifying profile-relationships, for which no mathematical solution is available, led to cutting crudely this Gordian knot by taking the modal or dominant characteristic as a single index.

May not the absence of any systematic relationship of moral behavior to social class found in this study but contradicted by other research be a function of restricted range, since the sample contained no cases above the upper middle class? Size of community may also have been a factor.

One wonders, too, about the effects on the 34 children and their social environment of the very frequently repeated interviewing, observation, and extensive testing. I was once told by a Nobel prize-winning physicist that the observed muzzle velocity of a bullet is a function of the method of measuring it. How much more may this be true of the kind of study here under review? Would not such a study possibly suffer from what survey researchers have called "panelitis," so that the sample does not remain representative of the population with respect to the characteristics under study? May such intensive study in a small community possibly produce the "Hawthorne effect" well known to industrial psychology? Obviously these latter are serious questions of methodology. Through future better research designs this methodology can be improved.

Having said all this, it is only fair to say that this research has broken new ground for further research in an area of highest importance if scientific truth is to replace folklore and folk wisdom in dealing with the complex problems of human intra- and inter-personal interaction and, even more importantly, inter-group behavior ultimately on a world scale.

Purdue University

H. H. REMMERS

Contemporary Theories and Systems in Psychology. By BENJAMIN B. WOLMAN. New York, Harper & Brothers, 1960. Pp. x, 613. $7.50.

Explanation of Human Behaviour. By FREDERICK V. SMITH. Second edition. London, Constable and Company, 1960. Pp. xi, 460. 30 s.

The person who tries to write a critical review of the main theories and systems in psychology is undertaking a very difficult task, one that calls for great breadth of knowledge, a very high degree of accuracy and critical skill, and an ability to understand sympathetically a diversity of conceptual systems based on quite different presuppositions. This ambitious task has been attempted by the present authors in two considerably different ways, with each of them making his major investments at different points, so that the two volumes well supplement each other.

Of the two, Wolman has discussed much the larger number of theorists. The chief merit of his book is the huge panorama which he paints, even though he has selected only three main areas for discussion—the areas of learning theory, psychoanalysis and related systems, and phenomenological and field theories. For example, in discussing conditioning, Wolman has not dealt solely with Pavlov and Bekhterev, but also with subsequent work in Russia and with Razran in this country; the discussion of Freud is rounded out by discussions of the ideas of Adler, Jung, Rank, Kardiner, Dollard and Miller, Horney, Fromm, Sullivan, and some others; the chapter on "personalistic psychology" is preceded by an excellent chapter on the philosophical and semi-philosophical backgrounds of the work of Spranger, Stern, and Allport.

There are some costs which we probably can charge to the hugeness of this field of discussion. (1) There are many rough spots in the book—places where there are inaccuracies, misstatements, or unnecessarily vague statements. It is probable that Wolman could not humanly have avoided most of these just by his own efforts, but it is unfortunate that a book like this could not have received a considerable investment of careful editorial work by experts in the various fields discussed: *e.g.* in the chapter on Freud, someone surely could have caught the mistake of entitling as "metapsychology" the section describing Freud's ideas about occult phenomena; Lashley ought not to have been described as a behaviorist in the main part of his career; Wolman should not have said that "Adler's main consideration was the way man relates to men in a conscious way" (p. 386). (2) Despite the title of his book, Wolman has made relatively little use of revisions and critical discussions of main theories in the last decade or so before the book was published. Later critical discussions are listed in the extensive bibliography, but generally were unused in the text. Lewin stops with Lewin, Hull with Hull, Gestalt thought with Koffka and Köhler. In fact, the most important later work of many main theorists is unmentioned, as in the case of Köhler's work on figural after-effects, Skinner's work on schedules of reinforcement, and Adler's shift of stress from physical deficiencies to interpersonal experiences on other scores as main sources of inferiority conflicts. (3) Some main concepts of some theorists have not been grasped, as in Wolman's describing Lewin's theory as over-environmentalistic and saying: "Only overt and manifest behavior . . . seems to be considered by Lewin" (p. 475). (4) The general evaluative comments are sometimes quite dubious (*e.g.* "Freud's personality model is constructed with the precision of a mathematician." (p. 354); "Hull's way of reasoning is impeccable" (p. 125). (5) Some discussions, rather than singling out main concepts for clear presentation, have tried to compress everything into a small space, as when Wolman uses a total of 7 pages to quote verbatim all of Hull's postulates and main corollaries. At such points, unless readers are familiar with what is discussed, the mass of new terms, symbols, and technical statements would be overwhelming.

Many of these difficulties, it may be hoped, can be remedied in a second edition. Woodworth achieved a much more excellent volume when he revised his *Contemporary Schools of Psychology* in 1948, and the present edition of Smith is a much more adequate book than his 1951 edition. Meanwhile, even though the unevenness of Wolman's book is serious, the book has real values.

In most of his chapters, Smith has attempted a more manageable task by taking a smaller number of key figures (G. Allport, Lewin, Freud, Watson, Hull, and Tolman). The six chapters on these men have a sort of "great books" approach; they achieve high clarity and interest, but are just on these men as such. A vastly different thing is attempted in the two chapters which deal with "the system of Wm. McDougall" and "the Gestalt approach." In each of these chapters, Smith does a really magnificent job of discussing a great quantity of related experimentation and theoretical discussion which has tested and extended the theories in question. The Gestalt chapter is limited to perception, rather than dealing also with learning, but its interrelating of theory and empirical evidence is splendid. The same is true of the discussion of recent neurological and ethological work which carries on from the discussion of McDougall's work, whereas Wolman's eight pages on McDougall leave him virtually as though he had failed to anticipate any crucial developments in psychology. It is regrettable that Smith could not have done the same sort of thing for each of his other theorists, but it would have been a superhuman job.

Both authors (Wolman in his last two chapters, Smith in his first five) discuss the general nature of explanation or theory-construction in psychology. Smith has some scattered observations in this section which are rather acute, but, in general, neither book is strong on this topic.

Whether any subsequent author can do better at the difficult task of writing a book on theories and systems in psychology remains to be seen. Maybe the task needs to be a distributed responsibility, as in the Koch volumes. Maybe a more adequate discussion could be obtained by a team drawn from different theoretical backgrounds but willing to work together and willing to let some one or two members of their group integrate what they provided. But the job looks like more than a single individual, unaided, can do.

University of Oregon

ROBERT W. LEEPER

Das Gestaltprinzip im Leben des Menschen und der Tiere. By KARL BÜHLER. Bern and Stuttgart, Verlag Hans Huber, 1960. Pp. 105. DM 22.

This book is important, not only because it suggests the warm and rich valedictory of a distinguished psychologist, but because it is an instructive assembly of data from many fields, directed toward an epistemological validation of the *Gestalt* principle and of its relation to the senses. The reader must be prepared for an exciting, though sometimes troubled and uncomfortable, canter through the origins of *Gestalt*, behaviorism, cybernetics, John von Neumann, Einstein, epistemology, novelties of neurology, language, robots, electronic computers, mythology, some Freud, social behavior, and varieties of psychophysics—all in 100 pages of text that is too discursive here and too terse there. With great ingenuity and propriety, data from the above subjects are surprisingly modulated into *Gestalt*.

For a work emphasizing *Gestalt*, it appears to have some ambiguities, and to cater to a much oversimplified image, when it deals with the origin of *Gestalt*. Here, I fear, lie traces of old polemical partiality and asperities. First, Bühler's hospitality is too prodigal to all but the actual founders of *Gestalt*. He gives priority to anyone, preferably ancient or Austrian, who may have casually observed that the sum of the parts is not the whole. This lets in Lao-Tse as a Founding-Father. This would not matter were it not for the fact that from the pre-Socratic philosophers through Bergson the ideograph of wholes was a stereotyped observation and nothing was done about it. Strangely he features the Austrian Von Ehrenfels as the actual founder of *Gestalt*, when obviously he contributed nothing basic. He merely shored up mechanistic elementarism while remaining unaware of the real problem, by adding secondary elements to primary elements. This is not *Gestalt,* as Bühler thinks it was. It remained for Wertheimer (he and Köhler are mentioned, but never Koffka, Gelb, or Goldstein) to initiate *Gestalt* by showing experimentally that there is no compounding of unaltered elements in a perceptual whole, and that parts are defined by relation to a system; and for Köhler to extend the evidence and to show that what holds for perception also holds for learning and for the external physical world. Koffka as a theorist and wide-ranging assembler, Gelb on constancies, and Gelb with Goldstein on neuropathological behavior are the next contributors in line. To all this Bühler is indifferent. "C'est magnifique, mais ce n'est pas . . ." the history of *Gestalt.*

Once done with this strange aberration of history, Bühler introduces some familiar and some startling considerations, but all relevant to the issues of the operation of the *Gestalt* principle in the life of man and animal and in the operation of the machine. In humans he rapidly presents the physical and mental development sequence for which he is known, as not from an infant homunculus to a large adult, but as from one kind of organism to a qualitatively different organism. In birds and other animals the qualitative difference may be less to zero, owing to the very short span of development. At all stages animals, birds, bees, and others are *Gestalt*-perceiving creatures. A spider fed on the blood of a human schizophrene undergoes progressive disorganization manifested in the progressive configurative dilapidation of his web-spinning: since the loss is the loss of a *Gestalt,* the original instinctive performance must itself have been configurative. Birds are fine celestial navigators, dependent on configurations of stars, though not necessarily on our traditional constellations. On the other hand, though robots can play a fair game of chess, they are unlikely to think, dream, or learn from experience; and electronic computers are also unlikely to become capable of creative thinking or of experiencing optical illusions, because they are destitute of *Gestalten.*

In all this remarkable effort to correlate machines with organisms, Bühler holds that the key lies in behavior determined by summation as against behavior determined by *Gestalten.* The criterion of organic life we may infer, then becomes the presence of physiological, together with psychological, *Gestalten.* Unlike organisms, whose performance is creative because configurated, machines, even when simulating human behavior like chess-playing, can never be creative because their acts are mere summations and never configurated performances. In other words, with the machine the whole is merely the sum of the parts. Moreover, the machine is handicapped configuratively by the limitations of its feed-back system of control, whereas the organism has the advantage of the flexibility of *gestalted* perception and behavior. This means that where a feed-back system would be stumped by a radical change in perspective of a visual stimulus, an animal could cope with it easily with the help of the visual constancies.

Perhaps the best summary of Bühler is that the *Gestalt* principle is a mysterious, vital, ordering force, recognizable in a three-fold manifestation of animal life:—nurture, growth, and performance (to which cognition belongs). By this principle our sensory experiences are related to one another and significantly tied with action. This relationship we may marvel over but never comprehend. As Einstein has written, "The eternally inconceivable of the world is its conceivability." To this paradox one could say Bühler has devoted his book.

Sarasota, Florida ELLIS FREEMAN

Mental Deficiency: The Changing Outlook. By ANN M. CLARKE and A. D. B. CLARKE (eds.). Glencoe, The Free Press, 1958. Pp. xvi, 513. $10.00.

This book, prepared by a half-dozen of the leading English psychologists working in the field of mental deficiency, treatments of mental deficiency almost exclusively from the psychological point of view. It is divided into three sections, dealing with the background, the theoretical problems, and practical problems in mental deficiency.

The first section of the book is primarily a treatment of the problem as it exists today in England. There is an extensive treatment of the laws concerning mental deficiency and the authors in several chapters seem to be unanimously opposed to the provision of the present law including the psychopathic with the mental defectives. While this has little relevance for American psychologists, the discussion is interesting and does bear on some problems encountered in this country.

The best part of the book is the section on theoretical problems. Here there is a thorough and scientifically sound treatment of many of the problems encountered in this field. Classification, measurement, genetics, individual differences, and learning are some of the main areas covered in this section. The material is sound and handled in a scholarly fashion. The review of the literature is very thorough. Although many of the references are primarily of the British literature, there is an amazingly complete coverage of the American literature as well. This is attested by the fact that there are over 700 references in the bibliography. Of some interest to American psychologists is the coverage given to recent Russian work in the chapter on "Learning and Mental Defect." Much of this will be of interest to American workers who probably have not had access to this literature. The discussion of brain damage and mental defect is particularly good because of its critical approach and evaluation of this whole area.

The section on practical problems will be of somewhat less interest to the experimental psychologist but parts of it may be of value to the clinical. The chapter on "Psychological Assessment in Mental Deficiency" will not contribute much to the practicing clinician who is working in this field but the student entering this area will find it quite valuable. Of interest in this section are the chapters on "Psychotherapy with the Feebleminded" because of its critical evaluation of therapeutic attempts, and the chapter on "Adoption and Fostering of Children of the Mental Deficient" because of its treatment of an area that has been largely neglected in the literature.

Generally, books written by a collection of contributors vary considerably in the quality of the writing. This book appears to be an exception. The editors apparently have done a very skillful job in bringing the various chapters up to the same high level of professional writing. In contrast to so many recent books in the field of mental deficiency, this book is intended for professional people and makes no concession to the lay reader.

The beginning graduate student or even the advanced undergraduate who intends to go into the field of mental deficiency will find this book extremely valuable. For the clinical psychologist who has worker in other areas but who may be moving into this field, or who may wish to know more about it, this book will be very useful in providing a theoretical orientation and an understanding of some of the practical problems in dealing with the mentally defective. While the orientation of the book is quite thoroughly psychological, workers in the educational area of mental deficiency will find Part 3 quite useful. This section will be particularly helpful to those working in public school settings, in vocational rehabilitation, and in various community programs in mental deficiency. For the experienced psychologist already working in these fields, this book will be useful in bringing him up to date on the English literature as well as some of the Continental work in this field. The references are quite thorough and seem to be quite recent up to the date of publication of this book.

Austin State School
Austin, Texas

WILLIAM SLOAN

Search for Security: An Ethno-Psychiatric Study of Rural Ghana. By M. J. FIELD. Evanston, Northwestern University Press, 1960. Pp. 478. $6.50.

Once before, the Northwestern University Series of African Studies contained a book of interest to psychologists—the Herskovits' *Dahomean Narrative.* In view of the excellence of that volume, one might be forgiven if disappointment is expressed regarding Field's treatment of pathological behavior in contemporary Ghana. On the other hand, we should be grateful to the Africanists for any material pertaining to behavioral abnormalities in rural, non-Western societies, for such material is sparse. Prevailing myths, such as the one that would have "less civilized" peoples subject to mental illness only as a response to the stresses of acculturation, are badly in need of the dispelling power of empirical fact. While Dr. Field's book might succeed in dispelling some such myths, it is likely, however, to contribute to others. Although this British psychiatrist attempts to indicate that not all Ghanaians develop pathological fears involving self-accusations of witchcraft (an eight-page chapter is entitled "Stable Personalities"), it is difficult to avoid that impression. For this is a rather sensational book, characterized by a generally unscientific brand of reporting.

Some 150 case reports, based primarily on observations of supplicants at Ashanti shrines, constitute the bulk of the book. To classify these cases, Dr. Field employed the standard European psychiatric categories which, she maintains, could in most instances be used "without any misgiving." To assume unquestionably the cross-cultural applicability of diagnostic labels which, even in their culture of origin, are considered suspect by many, is a serious error. How much more useful it would have been had Dr. Field utilized her observations in a partial test of the usefulness of the diagnostic categories *per se,* rather than simply forcing her cases into conformity with them. (It must also be said that this reader often found it difficult to detect the discriminating features that led to the classification of various cases into one category rather than into another.

Further reasons for questioning the scientific value of this book may be found in Dr. Field's many off-hand comments obviously based on hearsay, in the all-too-frequent snap judgments labeled "commentary," and in the repeated use of meaningless evaluations in which pre-morbid personalities are described as "good," "very good," etc. No examples will be given here; they may be found on almost any page.

In many ways Dr. Field's book is reminiscent of the missionaries' reports and travellers' diaries that once served as the major sources of information about non-Western cultures. The long-range value of those amateurish endeavors lay in their stimulation of skeptical curiosity which led eventually to the development of disciplined ethnographic reporting. Hopefully, books like Dr. Field's will stimulate a parallel development in the field of ethno-psychiatric research.

Until such development is forthcoming, a book like this will have to serve. It can, if the reader will take the trouble to penetrate the overly-personalized, anecdotal style of its author and concentrate instead on the examples of pathological behavior which she provides, often with admirable lucidity and richness of detail. Considered in the context of some of the more obvious facts of the social and cultural milieu of rural Ghana today, the case histories provide impetus to psychological theorizing.

One need not, after all, be swayed by the author's analyses, nor need not be concerned whether case 91 is or is not "a simple schizophrenia with three atypically short, acute episodes possibly precipitated by malaria." Most psychologists can and will draw their own conclusions. In so doing, those psychologists who have been relatively unacquainted with the nature of deviant behavior in societies other than our own may be forced to reconsider not only familiar diagnostic categories, but, more importantly, existing theories of the etiology of behavioral abnormalities. The African Studies may again have given psychologists something of value.

Columbia University MARSHALL H. SEGALL

The Child's Conception of Geometry. By JEAN PIAGET, BARBEL INHELDER, and ALINA SZEMINSKA. New York, Basic Books, Inc., 1960. Pp. vii, 411. $7.50.

This volume from Piaget's laboratory in Geneva is a companion piece to the *Child's Conception of Space* published in English in 1956. It deals primarily with the development of notions of measurement and geometrical concepts like coördinates, angles, areas, whereas the earlier volume was concerned more with the child's conception of the concept of space itself. Generally speaking, the present volume is concerned with later developments than the earlier one, although the division is not a clear one.

The discussion follows the now familiar pattern found in nearly all of the reports of empirical studies from the Piaget laboratory. Each chapter is concerned with some particular problem and reports two or three empirical procedures that were used to investigate the child's understanding of the concept. A sequence of two or three stages with occasionally some sub-stages is proposed to describe the development and several verbatim protocols are given to illustrate the sort of behavior found at each of the stages. Theoretical material is interwoven into the report of the empirical investigations.

In view of the extensive body of investigations on these topics, it is not surprising that some of the experiments reported in this volume are essentially confirmations of investigations previously reported. For example, the study of the child's conception of distance along a curved pathway is extremely similar to an earlier study reported in *Les notions de mouvement et de vitesse chez l'enfant.* These interlocking confirmatory studies are gradually building a solid structure of empirical findings, although the clinical concentric method that is the standard procedure in all of these investigations does not exclude the possibility that the findings can be contaminated by the hypotheses of the investigators.

This volume, as others, contains some real surprises in the sense that the child is discovered to lack certain tacit assumptions about the nature of space that are so ingrained that adults scarcely realize that there is any assumption involved. For example, in this volume, it is reported that the child does not believe that the distance between two points remains the same if objects are placed in the intervening distance. The assumption that distance is unaffected by the way that the intervening space is filled is, of course, a very basic one and it is quite important that young children fail to conceive of space as a category into which content may be fitted.

The investigations show very neatly the importance of the child's adopting a transferable unit as a standard against which distances, areas and volumes may be compared. This is an essential operation in the process of measurement. The studies also report some interesting investigations of the child's use of two coordinates as a procedure for locating a point in a plane. Tacit assumptions involved in such an operation are extremely interesting to investigate in young children.

So again, the psychological public is presented with a wide variety of very interesting observations about the development of children's conceptions. Judging from the results of repetitions of Piaget's earlier work, we can assume that the findings recorded in this volume are accurate descriptions of the behavior of some children in these situations. Children's behavior may not fit Piaget's theory quite so neatly and logically as they appear to from the reports in this volume, but once again Piaget and his colleagues have broken new ground and posed problems in developmental psychology upon which other research workers may feed for a number of years.

Cornell University ALFRED L. BALDWIN

Intelligence and Attainment Tests. By PHILIP E. VERNON. London, University of London Press, Ltd., 1960. New York, Philosophical Library, 1961. Pp. 207. $7.50.

Intelligence and Attainment Tests is a relatively short volume written primarily for students of psychology, teachers, and others interested in education, the social sciences, and guidance. Consequently, the book is written for the most part in a non-technical manner. For instance, little mention is made of statistical considerations in measurement, although a topic such as factor analysis is mentioned repeatedly, though briefly, throughout the entire book. The heaviest treatment of this topic occurs in the last chapter, and this section, like all previous remarks is non-mathematical in nature.

The book is divided into 11 chapters, followed by an appendix of 1 page and a bibliography of 8 pages. The latter contains many references to American periodicals and books. The book, however, is devoted primarily to descriptions of tests widely used in Great Britain.

The outstanding feature of Professor Vernon's book is the heavy emphasis on intelligence tests and the very light emphasis on attainment tests. One chapter is devoted to the nature of intelligence, another to individual intelligence tests, and a third to group intelligence tests. In addition, there is a chapter concerning the effects of coaching and practice on intelligence tests, the results of mental testing, and educational and vocational implications of intelligence tests. In contrast, only one of the 11 chapters is directly devoted to educational attainment tests; this chapter is a scant 11 pages in length.

In addition to the lack of emphasis on attainment tests, the reader will quickly notice the failure of Professor Vernon to devote sufficient space to validity, reliability, and norm considerations. The weakness of the treatment in these three areas probably stems in part from the fact that Professor Vernon attempted to restrict his consideration of statistical aspects of testing to a minimum. On the other hand, it is disappointing to note that Professor Vernon devoted considerably more pages to the question of the effects of coaching and practice on intelligence test results than he did to the determination and evaluation of the validity and reliability of those tests.

A relatively large part of the volume is devoted to descriptions of individual and group intelligence tests used in Great Britain. Little is provided here which cannot be found in test manuals or in standard references concerning commercially available tests. On the other hand, his treatment is concise and provides a convenient first step for a person interested in enlarging his knowledge of popular intelligence tests.

In summary, it is clear that the material prepared by Professor Vernon will, in general, be quite useful to the unsophisticated student of psychological testing. The more sophisticated reader will not, however, find this volume helpful in his efforts to enlarge his grasp of the present status of psychological testing, except in so far as it is applied in Great Britain.

Colorado State University J. STANLEY AHMANN

A System of Medical Hypnosis. By AINSLIE MEARES. Philadelphia and London, W. B. Saunders Company, 1960. Pp. xvii, 484. $10.00.

Dr. Meares' book may be considered a standard textbook in medical hypnosis. The psychologist, however, will be interested only in certain areas, especially Chapters 9 and 10 dealing with the "Atavistic Hypothesis of the Nature of Hypnosis." This hypothesis is an important contribution in the field.

Part I, "The Nature of Hypnosis," is well worthy of the consideration of the psychologist especially if he is interested in the abnormal field.

Part II, "The Induction of Hypnosis," will very naturally be of least interest; but Chapter 12, "The Emotional Relationship," and Chapter 19, "Moral and Ethical Aspects," deserve serious consideration.

Chapter 3, "Suggestive Therapy," contains much material which deals directly and indirectly with unconscious motivation, defense mechanisms, and throws very considerable light on the whole area of learning theory and the formation of personality. This part is largely devoted to the field of personality maladjustment.

Part IV, "Hypnoanalysis," will be of special interest to those psychologists active in the field of abnormal psychology. We point out that Chapter 48, "Hypnography," and Chapter 49, "Hypnoplasty," mark original contributions on the part of this author. Part V, "Hypnosis in General Medicine, ' will very naturally have little interest to the psychologist.

We regret that Dr. Meares does not discuss such recent developments as time-distortion and time-progression, two subjects which naturally grip the attention of the psychologist; also even more recent material dealing with the recovery of memories implanted while under deep chemical anesthesia. We congratulate him on his sane approach to the dangers of hypnosis, a matter which some of our friends in psychiatry are stressing to an extent that is somewhat tiresome.

We would strongly recommend that this book be included in the library of the psychologist whatever be his field of interest. Hypnosis sheds considerable light on personality formation and on learning theory, quite apart from its implications in the field of abnormal psychology.

Colgate University G. H. ESTABROOKS

The Unconscious before Freud. By LANCELOT L. WHYTE. New York, Basic Books, 1960. Pp. xiii, 219. $4.50.

While it is known that Freud did not originate the concept of the unconscious, Whyte attempts to go beyond such usually cited forerunners as Schopenhauer, Nietzche, Hartman, Herbart, or Leibnitz to show that even these predecessors but reflected and improved upon an idea that had existed for hundreds of years in Western Europe. He begins with two quotations from Plotinus in the third century: "Feelings can be present without awareness of them." "The absence of conscious perception is no proof of the absence of mental activity" (p. 79).

While St. Augustine, St. Thomas Acquinas, Shakespeare, and a dozen others are quoted prior to the seventeenth century, none of the quotations is any clearer than the two above from Plotinus. In the reviewer's opinion, most of these early quotations are only suggestive, and the concept of the unconscious does not come through clearly in Whyte's quotations until John Norris, one of Leibnitz's immediate predecessors, so that Whyte does not add as much to the usual accounts of Freud's forerunners as he believes that he does. After Norris, the quotations from Malebranche, Leibnitz, Kames, Herder, and Schiller become more convincing, and Schopenhauer and Nietzche clearly anticipate Freud at many points.

The Freudian scholar will find Whyte stimulating on the general point that Breuer and Freud had much more to build upon than the historically unsophisticated reader realizes, but he will not come away with a differentiated picture of exactly which elements were taken from predecessors by Freud and just what Freud himself contributed. For instance, the reviewer did not find evidence in the quotations that the concept of a *repressed* unconscious existed in any developed form before Freud, in contrast to the non-repressed unconscious which Whyte's survey concentrates on. Nor is there any mention of the following 10 aspects of the Freudian conception of the unconscious: (1) The idea of childhood repression (primal repression) as an essential pre-condition of adult repression. (2) The many specific defense mechanisms by which repression is achieved. (3) The role of anxiety in repression. (4) The concept of resistance. (5) The theory of the relation of symptoms to the unconscious. (6) The concept of unconsciousness as varying with the remoteness of differing types of material from ego recognition. (7) The role of identification and the internalization of anxiety-guilt in childhood and its relation to the repressed unconscious. (8) The relation of repressed unconscious to dream states. (9) The conception of unconscious mental processes following different laws (primary process laws) from conscious processes. (10) The frequent association of sex and hostility motives with unconsciousness.

Had Whyte been more differentiated about Freud's conception, he could have shown clearly what Freud took over and what he contributed. As it is, his book, having shown that the bare idea of the unconscious is centuries old, seems to dismiss Freud as merely one more figure in the long chain. He leaves the important historical task of assessing Freud's own contribution essentially untouched.

Princeton University PETER MADISON

An Introduction to Psychoanalytic Research. By KENNETH MARK COLBY. New York, Basic Books, Inc., 1960. Pp. ix, 117. $3.00.

Colby, a psychoanalyst, has listened carefully to the serious critics of psychoanalysis as a science. He believes that while the therapeutic situation is a rich source of data, psychoanalysis is at present open to criticism in several ways: for example, lack of accurate written records, lack of quantification, lack of follow-up. Colby proposes, however, that the tools of psychoanalysis be used to found a science, to formulate and test generalizations in the clinical setting within the framework of psychoanalytic concepts. He even goes so far as to suggest that statistical tests are needed in evaluating the dependability of the generalizations developed in this way.

Psychologists will be pleased to find that their repeated criticisms of the psychoanalytical method have affected at least one psychoanalyst, but they will also be disappointed in the progress made in this book. First of all the book is written at an exceedingly elementary level. If psychoanalysts must be approached at this low level of sophistication about the nature of science, one wonders what are the chances of their being able to adopt Colby's suggestions for research. Secondly, we are disappointed at the meagerness of the examples of research in psychoanalysis. As Colby says, "Undoubtedly many of the correlations thus far discovered in the psychoanalytic situation are trivial. In depressing fact, our established regularities are few and loosely stated and our causal explanations are still too global for precise prediction and modification of behavior" (p. 107).

In spite of the cautious evaluation expressed in the above quotation, the few examples which Colby does give seem to be accepted uncritically and even in violation of his own rules of evidence. One example is Freud's hypothesis that the traits of orderliness, stubborness and stinginess are correlated. According to Colby, "Subsequently this correlation has been confirmed by various psychological tests." (p. 89). He does not, however, consider Freud's hypothesis about the cause of this grouping to be established as yet. We are not told, however, why the evidence for the correlation is considered sound while the evidence for the relation to infantile sexual development is not considered adequate. In fact, we are not told the nature of the evidence in either case. In another example, "Through clinical observation a correlation has been noted between the symptom of impotence in men and their unconscious belief in coitus as a dangerous castration threat. Upon modification of this belief the impotence disappears. From this concomitant variation, a causal correlation has been formulated which is now a useful part of every analyst's knowledge" (p. 90). This reader is unable to discover in Colby's book any statement of how we observe unconscious beliefs, yet his example would seem to imply that they can be observed directly.

In short, Colby has sketched the skeleton of a sound approach to psychoanalytic research. Psychologists will be looking forward to the time when Colby and others can cover the bones with the flesh of observed regularities. T.A.R.

An Introduction to Psychoanalytic Theory of Motivation. By WALTER TOMAN. New York, Pergamom Press, 1960. Pp. x, 355. $9.00.

The writing of popularized accounts of psychoanalysis continues. Despite the title, Toman's book is mainly a general exposition of the hard-core of analytic thinking though, insofar as one might be willing to assume that psychoanalysis is mainly a motivational theory (the reviewer does not think it is), a case could be made for the title.

The first section of the book, dealing with concepts, begins with a discussion of psychological forces, for which Toman suggests "desire" as a generic label. Subsequent chapters in this section are then organized around the concept of *cathexis,* which "is the major 'operator' in psychoanalytic theory"; it is the mechanism that is responsible for determining what the vicissitudes of desires shall be. The second part of the volume, "The Theory at Work," is an account of personality development from the first year of life into the problems of the "choice of professions" and "marriage and parenthood." Three concluding chapters are devoted to psychopathology, psychotherapy, and a case of psychotherapy. With the exception of a few places the various topics are handled by the straight-forward application of classical Freudian concepts and analytic devices. This is particularly noticeable in the second part where the chapters on various developmental periods seem to rush along, an impression that is enhanced by the absence of chapter sub-divisions.

The volume inevitably brings to mind the writings of Theodore Reik; there is a distinct literary manner (not surprising, in light of the fact that Toman has published a novel, three books of short stories, and a volume of poetry). It is not a matter of literary style to which I refer, however; Toman resembles Reik in that he has a real-life instance, as all story-tellers have, for each point he wants to make. It makes interesting reading and is sure to appeal to the lay reader who is likely to be swept along most pleasantly through even the most formidable italicized terms and abstractions—there is certain to be another interesting instance on the next page. Unlike Reik, however, there are too many; Toman tries to make too many points that way and once it is realized that almost all the instances are actually *for-instances,* persuasively made up to suit the occasion, the reader immediately starts to make up counter-instances. In other words, there is little fact and much lore and argument. Toman does offer, however, standard psychoanalytic fare in recounting how the mind works and personality develops, he does it entertainingly, and there are places where an observation or illustration stands out by its shrewdness or aptness.

The book suffers because it does not have a clear public; it is written in a style suited to the layman, but it introduces too much that a layman cannot assess properly; on the other hand, it does not have enough depth to keep a professional psychologist busy for very long.

University of Oregon RICHARD A. LITTMAN

Bargaining and Group Decision Making: Experiments in Bilateral Monopoly. By SIDNEY SIEGEL and LAWRENCE E. FOURAKER. New York, N.Y., McGraw-Hill Book Co., 1960. Pp. x, 132. $4.90.

This book was the recipient of the Monograph Prize of the American Academy of Arts and Sciences for the year 1959 in the field of the social sciences. The authors, professors of psychology and economics, respectively, set out to test certain economic theories concerning bilateral monopoly (the situation in which a single buyer is confronted with a single seller) with crucial empirical data derived from carefully designed and executed experiments.

Starting with a chapter on theoretical formulations in which a mathematical model is proposed and the hypotheses to be tested are specified, the book describes in adequate detail the procedures of the experiments performed, the results of the experiments and ends with a chapter entitled "Overview and integration."

The experimental technique designed to test deductions from economic theory is simple and elegant. Pairs of college students were provided with isoprofit tables describing precisely various profit levels (in real money!) attainable and the price and quantities of Product X to be exchanged in order to reach a given level of mutual profit. The result of the bargaining between buyer and seller is that each goes home with the amount of money he decides is his largest possible profit. Bargaining is carried out in isolation by passing numerical price and quantity bids back and forth via the experimenter until the contestants are satisfied with the bargain achieved.

The hypotheses tested by these laboratory experiments centered around the case of equal bargaining strength, the theoretically indeterminate case. The main hypothesis, that contracts negotiated under simulated bilateral monopoly conditions will tend to the *quantity* which maximizes joint payoff, was confirmed. By varying cost and revenue functions, the generality of the result was confirmed. It was shown further that personal characteristics of the bargainers enter heavily into the determination of differential payoff and price. The variability of contrasts around the maximal joint payoff appears to be controlled by the amount of information the bargainers have. Results are reported for the complete-complete, incomplete-complete and the incomplete-incomplete information cases.

Finally, experiments are reported on the effect of manipulating the bargainer's expectancy by varying the amount of payoff in a "level of aspiration" design. Future experiments will explore the concepts of level of aspiration, panic behavior, interpersonal and individual personality and aptitude variables, sequences of failure and success experiences, forms of negotiation and the experimental method will be extended to duopoly and oligopoly models.

Princeton University

JOHN L. KENNEDY

Drugs and Behavior. Edited by LEONARD UHR and J. G. MILLER. New York and London, John Wiley and Sons, Inc., 1960. Pp. xix, 676. $10.75.

Sixty-two contributors wrote 53 sections and gave citations of more than 1400 authors. The book gives valuable information both to the specialist and the general reader. Some of the best portions describe methods of demonstrated or potential usefulness to drug studies, *e.g.* Cattell's lucid treatment of the measurement of anxiety and other mood reaction patterns; Vanderberg's eclectic review of methods for assessing neuroses and psychoses; Lorr's evaluation of nineteen rating scales and behavior inventories for assessing the status of psychiatric patients; Fleishmann's review of psychomotor tests; Kubie's stimulating discussion of drug research as viewed by the psychoanalyst; Nash's useful chapter on design and conduct of drug experiments (This reviewer's recent article on the same subject (*J. Psychol.*, 51, 1961, 173-211) differs sharply with Nash's assignment of so commanding a role for the logic of probability statistics). Many of the chapters have nuggets of insight and wisdom dispersed among otherwise less valuable material. Here and there one finds healthy scepticism of the shibboleths of mid-century psychological research. More than half of the book presents experimental results of drug studies. It is difficult to pass judgment on most of these since space restrictions apparently precluded presentation of enough evidence to make the data convincing. The spectrum is wide: from animal experiments to 'self-experiments' using free introspection.

Long sections of the book bored or irritated this reviewer. Some of the experiments described are trivial. One fails to understand the sustained enthusiasm of certain authors for finding negative results after administration of drugs of known effectiveness, *e.g.* effects of 800 mg. of meprobamate on simulated automobile driving. Equipment is described which has not yet been built, much less validated. The editors' call for more brass instruments for performance testing may aid a return to the unrewarding work of the period 1880-1950. Much is made of the need for 'objective' methods, but there is little clarity about the meaning of the word. (Reviewer's suggested definition: Objectivity is attained by a method when the rules governing its application are strictly defined. Objective methods can range from introspection to kymography, as can non-objective methods.) Too many of the researchers seem to be content with "placebos, double blind methods, adequate sampling techniques . . . and proper statistical evaluation of the significance of results" without appearing to be very much interested in what behavior systems are affected by a drug or what the meaning of their measurements is.

The printing and type face make for easy reading. The subject index does poorly with a most difficult task.

University of Rochester G. R. WENDT

A Rorschach Study of Child Development. By NETTLE H. LEDWITH. Pittsburgh, University of Pittsburgh Press, 1960. Pp. viii, 336. $6.50.

The author of this longitudinal study presents, in full detail, the Rorschach protocols of 11 children—6 girls and 5 boys (selected from a group of 138)—who were examined annually with this projective test, from the age of 6 to 11 yr., inclusive. These were all non-clinical cases; they were children who would be regarded as belonging in the 'normal' category. Yet, the selection was such that each child showed "unique and interesting variations in his developmental process" which attracted the interest of the author and her collaborators.

Each child's annual protocols were given 'blind' interpretations by one and the same psychologist; and a different psychologist was used for each child. These psychologists knew only the sex and Stanford-Binet *IQ* of each *S*. To validate the interpretations, the usual external criteria were used: interviews with parents and teachers, clinical observations by psychiatrists and "other psychological test data." In addition, however, since this was a longitudinal study covering a long period of crucial years in an individual's development, the author was able to provide the evidence from follow-up interviews with the *Ss* several years after the termination of the study. These interviews, the author states, "provide evidence of the predictive validity of the interpreters' comments on the direction the *Ss*' behaviors will take in the future." Since follow-up reports are given in addition to detailed protocols and blind interpretations, the reader of this book can draw his own conclusions regarding the Rorschach's validity in this instance. It appears to me that the author's statement is supported by the evidence. Ledwith states, furthermore, that while "transient impacts" produce some changes in behavior and attitudes, the child, in most instances, "seems to return to his more or less basic personality pattern observed in the very first record." This conclusion, too, appears to be warranted from the evidence given in the book.

Individuals disposed to disparage the Rorschach test should read this book carefully. They will not find the usual psychometric evidence of test reliability and validity; but they will find raw material and cautious interpretations that demonstrate clinical validity and usefulness of the Rorschach Test.

Cornell University FRANK S. FREEMAN

The Strong Vocational Interest Blank: Research and Uses. Edited by Wilbur L. Layton. (Minnesota Studies in Student Personnel Work No. 10). Minneapolis, University of Minnesota Press, 1960. Pp. v, 191. $3.50.

This book consists of a series of papers which were presented at a conference on the Strong Vocational Interest Blank. The conference was held at the University of Minnesota in 1955 in honor of E. K. Strong, Jr., a name that is almost synonymous with measurement of vocational interests. The participants were Berdie, Clark, Darley, Strong, Super, and Tyler, and a number of younger men.

The content of the papers is quite varied in subject and scope. Strong presented some findings from an eighteen-year longitudinal study of interests and his reflections on the use of the Strong Blank in counseling. Darley's paper was concerned with certain aspects of the theoretical basis of vocational interests. Berdie reviewed validation studies on the Strong Blank. Clark analyzed certain methodological problems in interest measurement. Super reported findings relative to the interest maturity scale from his study of career-patterns. Tyler discussed her program of research on the development of interests. The other papers, presumably, were based on dissertations completed at the University of Minnesota.

The psychologist who has kept up with the literature on interest measurement will find very little new information in this book. Many of the findings on which these papers were based have been published in other sources. The reader will, nevertheless, probably be interested in the somewhat informal observations and comments by these 'pros' which are not usually included in published technical reports. Strong's reflections on the use of the Strong Blank will be of particular interest.

University of California, Berkeley LAWRENCE H. STEWART

Becoming More Civilized: A Psychological Exploration. Leonard Doob. New Haven, Yale University Press, 1960. Pp. xii, 333. $6.00.

"This is a psychological exploration, not a handbook or an apocalypse." From the first page to the last this work is characterized by restraint, caution, and true scientific humility. The problem that the author assigned himself is timely and important when we consider that underdeveloped countries are among the most significant current questions of the day. Doob argues that we might make discoveries about civilization from observing peoples as they become "more civilized." Two principal questions follow from this formulation: (1) Why do people in less civilized societies become more civilized in certain aspects? (2) What happens to them as they become more civilized? These questions may lead to generalizations despite the acknowledged fact that each society is unique. The data of the study derived from the author's own research in Africa and Jamaica and from the work of others in Africa and elsewhere. The plan of the book centers around the presentation of 27 hypotheses summarizing knowledge concerning social contact for acculturation. The problems from which the data were derived are not grandiose, the author preferring smaller, more manageable tasks that would yield quantitative results. In the reviewer's opinion, Doob has done a solid, substantial piece of work. Within the larger cultural or social framework he has managed to stick to his last in focusing upon the essential psychological data and has derived significant results. Admittedly such an approach has not been frequently attempted.

University of Wichita N. H. PRONKO

Americans View Their Mental Health: A Nationwide Interview Survey. By GERALD GURIN, JOSEPH VEROFF, and SHEILA FELD. New York, Basic Books, 1960. Pp. xxxvi, 444. $7.50.

This book is the fourth in a series published by the Joint Commission on Mental Illness and Health. It reports the results of an interview survey of 2460 adult Americans undertaken by the University of Michigan's Survey Research Center. There are 10 other volumes in the Joint Commission series, including the final report, *Action for Mental Health.*

Essentially, the report consists of tabulations of and commentary on responses to questions regarding feelings of gratification, satisfaction and aspirations. Feelings of distress, worry, and tension are evaluated. The actions of the interviewee in trying to solve his personal problems are explored with emphasis on the seeking of help from external sources. The authors are concerned with the differences in response characteristics associated with demographic class membership.

For most psychologists, the book would be tedious reading indeed. The authors appeared to be somewhat uncertain as to their potential audience and compromised by presenting material at a level somewhat below that of the professional psychologist or statistician both in the discursive and tabular sections. The layman, however, is not likely to find the volume interesting as there is little appeal in the rather monotonously written pages of discussion.

University of Minnesota EDWARD S. SULZER

Style in Language. Edited by THOMAS A. SEBEOK. Cambridge, Mass., The Technology Press of Massachusetts Institute of Technology, and New York and London, John Wiley & Sons, 1960. Pp. xvii, 470, $9.50.

This book is a collection of papers with comments presented at a conference on style in language sponsored by the Social Science Research Council and held at Indiana University in 1958. The conference brought together linguists, anthropologists, psychologists and literary critics to discuss the nature and characteristics of literary style. The psychological approach was represented by the following papers: "The Pronouns of Power and Solidarity" by Roger Brown and Albert Gilman, "Vectors of Prose Style" by John B. Carroll, "Some Effects of Motivation on Style of Encoding" by Charles E. Osgood, and "Commonality of Association as an Indicator of More General Patterns of Verbal Behavior" by James J. Jenkins. George A. Miller and Roger Brown represented the psychological point of view in a section containing comments relevant to the papers and the conference as a whole. Much, but not all, of the empirical work and points of view in the psychological area have been published separately in journals. The rest of the book contains a variety of linguistic and literary approaches of only minimal interest to the general psychologist.

Cornell University GEORGE J. SUCI

Working with Groups: Group Process and Individual Growth. By WALTER M. LIFTON. New York, John Wiley and Sons, 1961. Pp. ix, 238. $6.00.

This is an unfortunate book. It seems to be intended mainly for students in colleges of education who are planning to be guidance counselors; but the author hopes it will be useful to almost anyone working with groups. The essential difficulty is that Lifton never really analyzes the requirements of the different social settings in which group work is practiced, and as a result he encourages the unwary student to transfer the goals and methods of adult group psychotherapy to the ordinary public school classroom. In case the student doubts the legitimacy of this transfer, he is told: "It is possible to use terms like teaching, counseling, or psychotherapy interchangeably without doing violence to the kinds of relationships that need to be developed in a group to achieve the goal of individual growth" (p. 21). Lifton's major example of therapeutic group work in action is a course which he taught for academic credit at the University of Illinois.

This reviewer believes that this volume will do more harm than good in the circles to which it is addressed. Its main value lies in the very comprehensive bibliographical references at the ends of the chapters, which provide an excellent guide to the scattered literature of a rather confused field.

Cornell University JOHN HARDING

Hypnosis in Treatment. By WILLIAM MOODIE. New York, Emerson Books, 1960. Pp. 168. $4.00.

This book contains 23 chapters, 168 pages, and is concerned with such problems as preliminary interviews, hypnotic sessions, hypnosis and academic failure, nasal catarrh, hypnoanalysis, colitis, emotional maturity, behavior disorders in children, and so on. Simple arithmetic suggests that not much more than seven or eight pages can be devoted to each of the book's many topics. This book is intended for the doctor who "understands and is in sympathy with people." To this one might say "Amen" but might wonder who would be fool enough not to include himself. The main question that occurs to one throughout this text concerns the efficacy of suggestion versus suggestion-in-hypnosis. Moodie answers this question by assuming, although empirical evidence is lacking, that emotional release and less accompanying tension is more characteristic of the latter. One also encounters many puzzling statements concerning the mechanism of enuresis, hypnotic clearing of mental vision, who make the best hypnotists, the concept of 'will,' unconscious resistance, etc. The reviewer has little to say about this book for the simple reason that this book has little to say.

Washington State University F. L. MARCUSE

Experimental Psychology: A Methodological Approach. By F. J. MCGUIGAN. Englewood Cliffs, N.J., Prentice Hall, 1960. Pp. iv, 314. $6.00.

This introductory text describes basic experimental methods and the statistical analysis appropriate for a number of experimental designs. The author attempts to present enough philosophy of science to give the reader insight into the design, analysis, and interpretation of psychological research.

The text goes well beyond many introductory experimental texts both in terms of research designs and the statistical analyses. The reader is introduced to analysis of variance for designs including a 2^3 factorial design. Of special note is a detailed presentation of the Duncan's Range-Test, including 1% and 5% significance tables and an extension of the test to groups of unequal *n*.

The usefulness of the text may be limited in that it treats both experimental methods and statistical analysis in an integrated fashion. This approach is, in the reviewer's opinion, an excellent one but is at odds with the traditional separation of the two in many universities. The book is to be highly recommended for those psychologists who have minimal training in experimental design and statistics.

Washington University DANIEL L. LORDAHL

THE AMERICAN
JOURNAL OF PSYCHOLOGY

Founded in 1887 by G. STANLEY HALL

Vol. LXXV SEPTEMBER, 1962 No. 3

ITEM-SELECTION IN PAIRED-ASSOCIATE LEARNING

By BENTON J. UNDERWOOD, ROBERT REHULA, and
GEOFFREY KEPPEL, Northwestern University

The learning of a list of paired associates may be divided logically into two stages. The first stage is the response-learning stage during which *S* must acquire the response-terms so that they are readily available in recall. The second stage is the associative stage during which an association is formed between the stimulus- and the response-term. Some recent studies purport to demonstrate that the second or associative stage takes place in a single trial; trials preceding the one on which the association is formed are said to be irrelevant to its formation.[1] This conception is opposed to one which would assert that there is a gradual increase in associative strength between the stimulus- and the response-term as a function of trials, an increase which usually occurs both before and after the first correct anticipation is given.

The present paper is not directly concerned with whether or not an association is established in one trial. We do not, in fact, find the notion of one-trial learning discomforting provided certain additions are made so that other facts may be handled. For example, when *Ss* learn a standard paired-associate list consisting of eight pairs with single-digit numbers as stimuli and single letters as responses, the probability of an item's being correct on the trial immediately after the trial on which it was first given correctly is 0.65.[2] In such a list, intralist similarity is essentially zero, and

---

* Received for publication January 23, 1961. These studies were supported in part by Contract Nonr-1228(15) between Northwestern University and the Office of Naval Research.

[1] Irvin Rock, The role of repetition in associative learning, this JOURNAL, 70, 1957, 186-193; Irvin Rock and Walter Heimer, Further evidence of one-trial associative learning, this JOURNAL, 72, 1959, 1-16.

[2] B. J. Underwood and R. W. Shulz, Studies of distributed practice: XXI. The effect of interference from language habits, *J. exp. Psychol.*, 62, 1961, 571-575.

response-learning which is required is minimal. If one-trial learning occurs, this probability should be 1.0 or nearly so; thus, unless the notion of one-trial learning is supplemented with other ideas about learning or performance to account for such findings—of which there are many in the literature—very little progress has been made. Many such facts seem to have been overlooked by those who espouse the notion of one-trial learning. Still, it is not the purpose of the paper to attack one-trial learning; rather, we, like others, are concerned with the experimental techniques which have led to the presumption of one-trial learning.[3] Already the idea of one-trial learning of verbal material is being accepted uncritically by some.[4] If this acceptance stems only from experimental results, and if those results are based on faulty techniques, then clearly a point of method needs careful examination.

The original experiment leading to the inference of one-trial learning is relatively simple.[5] Twenty-five single letters (all except I) and 25 corresponding double letters (*e.g.* AA) were paired randomly with the numbers 1 through 50. The letters were the stimulus-terms, the numbers the response-terms. From this pool of 50 pairs, 12 pairs were drawn randomly to form the first list and, after replacement, a second list was drawn from the pool of 50, and so on. Members of the control group simply learned such lists to a criterion of one perfect recitation. Members of the experimental group started with the same lists that were learned by members of the control group. Pairs for which *S* gave an incorrect response, or no response, were, however, eliminated at the end of the trial, and new pairs (from the remaining pool of 38) were used as replacements so that the list always contained 12 pairs. The method of learning was that of alternate study- and test-trials; all items missed on any test-trial were replaced with new items for the next study-trial. The *S*s in the experimental group were, like those in the control group, required to achieve a criterion of one errorless test-trial.

The results showed that the mean number of trials required to achieve the criterion did not differ for the two groups. It was argued, therefore, that the experimental *S*s must not have developed any associative strength between the stimulus- and response-terms for the items missed. Had such associative strength developed, the elimination of the missed pairs should have retarded over-all learning; for the new pairs the associative strength would have to be built up from 'scratch.' The fact that the insertion of new items did not slow down learning is taken to mean that no strength was built up for the missed pairs and, consequently, that association must be an all-or-none affair.

Most other investigators have been able to confirm the original findings when the letter-number pairs have been used; therefore, the basic experimental facts do not seem to be in question.[6] Comparable findings using nonsense-syllable pairs have been seriously called into question by Postman's experiments.[7] Nonsense-syllables, however, probably require considerable response-learning; hence more is involved than the associative stage of learning. The one-trial learning case for the letter-number pairs seems intact since response-learning which is required is minimal for such pairs.

---

[3] L. L. Clark, T. G. Lansford, and K. M. Dallenbach, Repetition and associative learning, this JOURNAL, 73, 1960, 22-40; Leo Postman, Repetition and paired-associate learning, this JOURNAL, 75, 1962, 372-389.

[4] G. A. Miller, Eugene Galanter, and K. H. Pribram, *Plans and the Structure of Behavior,* 1960, 137.

[5] Rock, *op. cit.,* 187-188.

The obvious question to raise concerning the procedures followed throughout the above mentioned studies in which letter-number pairs were used is whether the insertion of new pairs (and the dropping of missed pairs) for the experimental *S*s may not have produced a factor (or factors) which facilitated learning. Thus, if one assumes a gradual buildup of associative strength, the negative effects produced by the withdrawal of items (having some associative strength but not enough to produce correct anticipation) might be counteracted by a positive effect resulting from the insertion of new items. It is, of course, disconcerting and unconvincing to have to argue that the positive and negative factors were in balance so that the net effect resulted in a rate of learning corresponding to that of the control group. Nevertheless, we will attempt to show that a positive factor can be identified experimentally.

What positive effect could be produced by the insertion of new items for the ones missed? The most prominent possibility is change in item-difficulty; the pairs used to replace pairs which are missed may be easier to learn than those that are replaced. Differences in ease of learning of pairs may be due to an already existing associative connection between the stimulus- and response-terms, or in the ease in which such associative strength may develop, whether in one trial or in several trials. This factor has been recognized as a possible contaminating influence by others (including Rock, who performed the original experiment), but attempts to evaluate its effect, as Postman has pointed out, have been inappropriate.[8] The evaluation can be made directly by adding a second control-group. The *S*s in this second control-group would learn the lists which constituted the final lists for the experimental *S*s. If there has been a selection of easier items as a consequence of the procedures given the experimental group, the learning of the lists by the second control-group should take place more rapidly than that of the original control-group. Of course, if there is no inter-*S* agreement on easy and difficult items—if selection is entirely idiosyncratic—the suggested new control-group would not demonstrate this selection.

---

[6] Clark, Lansford, and Dallenbach, *op. cit.*, 22-40; Michael Wogan and R. H. Waters, The role of repetition in learning, this JOURNAL, 72, 1959, 612-613.

[7] Postman, *op. cit.*, 379 ff.

[8] Postman, *op. cit.*, 375 f.

Some investigators who have worked with the letter-number pairs have reported that there is no commonality among *S*s as to easy and difficult items.[9] Our experiences with various types of learning material in which there *has* been high consistency among *S*s concerning easy and difficult pairs led us to doubt that the letter-number pairs would be an exception. In fact, one of the outcomes of the present studies is a demonstration that there are consistent differences in the difficulty of the letter-number pairs. Concomitantly, we were able to evaluate the amount of selection which occurred when missed items are replaced with new items during the learning of the *S*s in the experimental condition.

Five major experiments are reported, along with two minor ones. The first experiment was designed merely to show that selection of items will occur in a predictable manner when the items are of known difficulty. This experiment deals with pairs in which large differences in amount of response-learning are required; as such, the findings will not call into question any conclusions concerning the second stage of learning, *i.e.* the establishment of the association between the stimulus- and response-term. All the other experiments in this report make use of the letter-number pairs; from these we will attempt also to demonstrate that predictable selection occurs.

EXPERIMENT I

Materials. A total of 44 pairs made up the pool of items. The stimulus-terms for the pairs were the numbers 1 through 44. The responses consisted of two-letter units (bigrams) of which 22 had low associative connection between the two letters making up each bigram, and 22 had high associative connection. The degree of association was determined from tables of association.[10] A high-association bigram is illustrated by 'GO,' a low-association bigram by 'RZ.' The 44 numbers and 44 bigrams were paired through the use of a table of random numbers. Each of the 44 pairs was typed in the center of a 3 × 5-in. card, with only the stimulus-term typed on the reverse side.

Three groups of 30 *S*s each were used. The procedure for two of these groups was essentially the same as that used by other investigators, while the third group is a control group needed to evaluate selection.

Control 1 (C_1). Through the use of random numbers, 30 different orders of the 44 pairs were constructed. The first 12 pairs in each of the 30 orders constituted the 30 different lists learned by the *S*s in C_1. The pairs were presented by use of the cards on study-trials; test-trials were given simply by using the reverse sides of the cards on which only the stimulus-terms were typed. Learning was carried until one perfect recitation was achieved. On the study-trials each pair was shown for 3 sec., with 5 sec. between successive pairs. Following each study-trial a test-trial was given on which each stimulus-term was shown for 5 sec., with *S* instructed to try to give the appropriate response-term. To be counted correct the response had to be given within the 5-sec. interval. Between successive cards on the test-trials, 3 sec. elapsed. Five different random orders of presenting the 12 cards were used on study- and test-trials. The *S*s were never shown nor informed of the correct responses on test-trials. During the learning trials *S*s were required to say aloud both the stimulus- and response-terms, and on the test-trials were required to say the stimulus-term aloud whether followed by the correct anticipation of the response or not.

---

[9] Clark, Lansford, and Dallenbach, *op. cit.*, 26.

[10] Underwood and Schulz, *Meaningfulness and Verbal Learning*, 1960, Appendix F.

Drop-out condition. This condition has been called the experimental condition in other investigations. Since it will be necessary to refer to different experiments, some confusion may be avoided, however, by calling it the 'drop-out condition' (*DOC*). Each *S* in this group was initially given one of the 30 different 12-pair lists; thus, these initial lists were exactly the same as those presented the *S*s in C_1. All pairs missed (wrong response or no response) on the first test-trial were, however, removed after this trial, and new cards were inserted so that for the second study-trial there were still 12 pairs. Any items missed on the second test-trial were replaced with new ones for the third study-trial, and so on. The order in which cards were used as replacements followed the random orders noted earlier; necessarily the order differed for each *S*. If the pool of items was exhausted before *S* achieved the criterion of one perfect recitation, replacements consisted of the items which *S* had first missed. Instructions given before learning told *S* that the items might change from trial to trial, these same instructions being given to the two control groups.

For the *DOC*, 30 sec. were required after each test-trial to replace missed items. A corresponding period of 30 sec. was inserted after each test-trial for the *S*s in C_1 and C_2. For all three groups the interval between study- and test-trials was 15 sec.

Control 2 (C_2). These *S*s learned a 12-pair list in exactly the same manner as the *S*s in C_1. However, the 12 items learned by *S*s in this group consisted of the 12 items with which some member of the *DOC* had achieved the criterion. Thus, each member of C_2 learned a different 12-item list, the particular list being determined by an *S* in *DOC*. If there is selection occurring in the *DOC*, the lists given the members of C_2 should be mastered more rapidly than should the lists given under either C_1 or *DOC*.

Of necessity, an *S* in C_2 had to follow an *S* in *DOC*. Thus, the *S*s were run in 30 successive blocks of three *S*s each. On a chance basis, the first *S* in a block was assigned either to C_1 or *DOC*. If C_1 came first, the second *S* was placed in *DOC*, with the third *S* in C_2. If the first *S* in a block fell in *DOC*, the second *S* was, by chance, placed either in C_2 or C_1, with the third *S* in the remaining condition. This assignment of the *S*s makes it highly probable that the three groups did not differ appreciably in learning ability.

Results. Fig. 1 shows the acquisition-curves for the three conditions, the curves being plotted in terms of mean number of trials to reach criteria 1 through 12 correct on a single trial. The *Ss* in *DOC* required the greatest number of trials to attain the criterion of one perfect trial (7.27 trials), C_1 the next greatest (5.77), and C_2 the least number of trials (4.30). The *F* of 9.88 is significant far beyond the 1% level. The difference between any two means exceeds the 5% level of significance as evaluated by *t*. It is concluded that the differences in conditions produced differences in rate of learning. The lists constituted of the 12 items appearing on the last trial in *DOC*—the lists learned by members of C_2—are clearly easier lists than those learned by *Ss* in C_1. This suggests, then, that a selection process took place during the learning of the lists by the *DOC*. Further evidence on the nature of this selection is presented below.

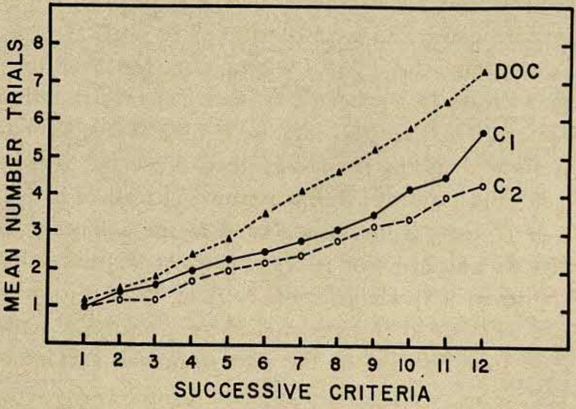


FIG. 1. ACQUISITION CURVES FOR EXPERIMENT I

Measures of item-difficulty may be obtained from the learning records of *Ss* in C_1. The first measure calculated was the proportion of times items with responses of high-associative connection (*H*-items) were given correctly, and the proportion of times items having responses of low-associative connection (*L*-items) were given correctly. Thus, if there were 5 *H*-items in a list and *S* required 6 trials to learn, there would be a total of 30 opportunities to give these *H*-items correctly. The proportion of times *S* did in fact give the items correctly was determined. The same proportion was calculated for the *L*-items. The mean proportion for the *H*-items for the 30 *Ss* was 0.74, and for the *L*-items, 0.61. The *t* for the difference between the means is 6.79.

As a second measure of difficulty of *each* item, we determined the proportion of times each was given correctly in C_1. On the average, for the 30 *S*s, each of the 44 items would be expected to occur in 8 different lists. In actual fact the number of occurrences varied from 2 to 13 times. Therefore, the values for each item have different degrees of stability. Nevertheless, when a mean proportion correct is obtained for each item and used to calculate a point-biserial correlation with *H*- and *L*-items as the other variable, the correlation is 0.57 (significant beyond 1% level). It is clear that the expected differences in learning *H*- and *L*-items occur in C_1. We may now ask, therefore, about item-selection during learning under *DOC*.

The 30 random lists initially given to the *S*s in *DOC* were constituted of 46% *H*-items and 54% *L*-items. The lists on which the *S*s in this group achieved the criterion were constituted of 64% *H*-items and 36% *L*-items. Clearly, there was a heavy selection of *H*-items as the *S*s learned. The average increase in number of *H*-items from Trial 1 to the end of learning was 2.1. The *t* for this increase is 6.42. Three *S*s showed no increase, two showed a loss of one *H*-item, while the remaining 25 *S*s showed a gain of at least one, with a maximal increase of five. The amount of increase in *H*-items is inversely related to the number at the start of learning. Five *S*s who had 4 or fewer *H*-items at the start gained an average of 3.8 *H*-items; 9 *S*s who had 5 *H*-items at the start gained an average of 3.8; 10 *S*s who had 6 *H*-items at the start gained 1.4 items, and 6 *S*s who had 7 or more *H*-items at the start gained an average of 0.7 such items.

The above data have shown that *H*-items are easier to learn than *L*-items, and that the drop-out procedure results in an increase in the number of *H*-items. The increase in number of *H*-items is associated with a greater ease in learning by *S*s in C_2 than by *S*s in C_1. It is presumed that the greater frequency of *H*-items in C_2 is responsible for this greater ease of learning.

Discussion. Experiment I has shown that a rather heavy selection of easy items occurs in learning by the *S*s in *DOC*. This heavy selection is undoubtedly due largely to the differences in difficulty in acquiring responses as such, since the items varied in terms of the strength of the initial association between the two letters making up the responses. Therefore, the data cannot be considered as evidence against one-trial learning; but they do indicate that item-selection can occur when the 'drop-out method' is used, and this demonstration was the purpose of the experiment. We shall now turn our attention to experiments in which response-learning is minimal, and where the emphasis is, therefore, on the association between the stimulus- and the response-terms.

EXPERIMENT II

(*A*) In this experiment, and in all following experiments, we have used a 50-pair pool of items consisting of the material originally used by Rock in his experiments from which one-trial learning was inferred.[11] The stimulus-terms were all single and double letters (excepting I), and the response-terms were numbers 1 through 50. The letters and numbers were paired through use of a table of random numbers, and 30 different orders of the 50 pairs were obtained, again using random numbers.

Procedure. Most of the details of Experiment II *A* were the same as for Experiment I, and essentially constitute a replication of Rock's original study. The three conditions were C_1, *DOC,* and C_2, with 30 *S*s in each condition. In presenting the pairs on study-trials, and the stimuli on test-trials, the time-factors were the same as in Experiment I. The *S*s, however, were not required to pronounce aloud the stimulus-terms on the test-trial, nor the stimulus- and response-terms on study-trials.

TABLE I

EXPERIMENT IIA: MEAN NUMBER OF TRIALS TO LEARN

	Mean	*SD*
Control 1 (C_1)	3.80	1.47
Drop-out Condition (*DOC*)	4.40	1.47
Control 2 (C_2)	3.40	1.28

Results. The mean number of trials to learn under each condition is shown in Table I. The *F* for the three distributions is 3.69. With 2 and 87 *df,* the 5% level of significance is 3.11. When differences are evaluated by *t*-tests based upon the within-groups variance, only the difference between C_2 and *DOC* is significant, the *t* being 2.70.

Previous investigators have found the performance of C_1 and *DOC* to be almost identical; for the present data the *t* for the difference between these two means is 1.62. Thus, while the difference is somewhat larger than that found by others, statistically speaking it must be concluded that the present results confirm previous studies. Looking ahead, however, it may be noted that in three additional studies the learning under *DOC* has always been inferior to the learning under C_1.

As noted earlier, if item-selection occurs as a consequence of the procedures given the *DOC* *S*s, C_2 should be superior to either C_1 or *DOC*. The differences are in the appropriate direction but only the difference between C_2 and *DOC* meets acceptable statistical standards. Therefore, the case for item-selection has not been demonstrated convincingly for these materials. We turn, therefore, to other methods of analysis.

---

[11] Rock, *op. cit.,* 187.

On the average, only 15 new cards were introduced into the lists for the *DOC Ss*; hence, for the average *S* nearly half the items were not subjected to possible selection. Nevertheless, it will be worthwhile to see if there is any inter-*S* consistency in the items 'selected' by the *DOC Ss*. The number of times each item appeared in *DOC* (including the items making up the initial list as well as those substituted for replaced items) was first determined, followed by a counting of the number of times each item appeared in the final list of this group. Then, for each item, we calculated the proportion of times an item appeared in the final list per number of opportunities to appear. The assumption is that those items with the highest proportions are the easiest items. As a second measure of item-difficulty the records of the *S*s of C_1 were used. From these we obtained a proportion correct per opportunity for each pair just as was done for the items used in Experiment I. For the 50 pairs, the product-moment correlation between the learning and the selection-measure for *DOC* was 0.37. With 48 *df* this value is beyond the 1% level of significance. It thus appears that there is consistency among *S*s concerning easy and difficult items in these letter-number pairings. We now present two minor experiments dealing with this matter of item-selection.

(*B*) Two lists consisting of eight items each were selected. One group of items was selected to form an easy list based on the selection-index derived from *DOC Ss*, and the other list formed a difficult list based on the same criterion. This index, as explained above, represents the proportion of times an item appeared in the final list of *DOC* per number of times it appeared on any trial. The eight easy items (List 1) based on this index were: RR-41; SS-45; J-16; U-50; S-5; BB-17; XX-11; and C-2. The difficult list (List 2) consisted of the following eight items: O-22; VV-15; JJ-29; AA-43; WW-18; Q-14; QQ-9; and ZZ-4. In addition a third list of eight items was used. This list was used by Clark, Lansford, and Dallenbach, and consisted of the following pairs: T-39; E-32; LL-38; JJ-23; PP-30; L-49; P-27; and Z-37.[12] The particular reasons for including this list may be better explained at a later point, so only the procedure and results will be presented here.

A total of 36 *S*s learned all three lists in counterbalanced order. The lists were presented on a memory-drum at a 2:2-sec. rate under the usual anticipation-learning procedure. Each list was presented for 10 trials with *S* instructed to get as many correct on each trial as possible. Four different orders of presenting each list were employed.

Results. The response-measure was the mean number of correct anticipations given in 10 trials. Since there was no interaction between order and lists, order was disregarded. The mean numbers of correct anticipations were 68.69, 54.75, and

---

[12] Clark, Lansford, and Dallenbach, *op. cit.*, 35.

34.86, for Lists 1 through 3, respectively. All differences between means are significant beyond the 1% level. The three acquisition-curves were clearly separated from the first trial through the tenth. Thus, the list made up of items judged to be easy as determined by the selection-index derived from *DOC* of Experiment II A was learned much more rapidly than the list determined to be difficult by the same index. In addition, the list used by Clark, Lansford and Dallenbach appears to be much more difficult than a difficult list derived from the selection-index of Experiment II A.

The results for Lists 1 and 2 would seem to support the notion that there is clear selection in the *DOC* of Experiment II A. The interpretation, however, is not completely clear. A count of errors showed significantly more overt errors per trial occurring in learning the difficult list than in learning the easy one. Thus, although inspection does not indicate clear differences in intralist similarity, the differences in errors indicate that this is a real possibility. Differences in learning, therefore, could be due to intralist similarity rather than to differences in ease of associating particular stimuli with particular responses. To clarify this matter, Experiment II C was undertaken.

(*C*) A single list was constructed to sample the entire range of difficulty of items as determined by the selection-index based on *DOC* of Experiment II A. The lists consisted of the following 10 pairs: O-22; XX-11; G-32; MM-6; F-20; ZZ-4; DD-25; H-1; VV-15; W-18; C-2; and BB-17. In this list double and single letters are used equally often and no letter occurs both as a single and double letter. Furthermore, there is minimal duplication among numbers. In short, there is no reason to believe that intralist similarity would produce differential disturbance in the learning of the pairs. The list was presented for 10 trials on a memory-drum exactly as in Experiment II B. Sixteen *S*s were used.

Results. For each pair the total number of correct responses was determined across all 16 *S*s. These values were correlated with the selection-index for each item as derived from *DOC* of Experiment II A. The product-moment correlation was .79, significant beyond the 1% level. We conclude from Experiments II B and II C that the selection-index derived from Experiment II A is a reliable one; there are consistent differences in the ease of learning the letter-number pairs.

EXPERIMENT III

The results of Experiments II A, B, and C indicate that there is item-selection occurring during the learning of the lists by the *DOC* *S*s, and that this selection is a predictable one. That is to say, items difficult for one *S* are likely to be difficult for another *S*. The puzzle that remained was why, if there is consistent item-selection, the *S*s in C_2 did not learn more rapidly than those in C_1 in Experiment II A.

The procedure used in Experiment II A produced very precipitous learning when that learning was measured in terms of trials. The slow rate of presentation and the 30 sec. between successive trials during which time most *S*s obviously rehearsed, are probably responsible for this rapid acquisition. It seemed possible that if the number of trials was increased, the amount of selection which would occur would be greater (more easy items selected per *S*) than that observed in Experiment II A. This could probably be accomplished by any or all of at least three changes in procedure; namely, increasing the rate of presentation, controlling rehearsal between trials, or increasing the number of items in the list. Finally, some concern was felt about the manual presentation of items which, while becoming fairly accurately timed with practice, might still be subject to subtle biases. Experiments III, IV, and V were designed to investigate the effect of such factors on the selection in *DOC*.

Procedure. Experiment III introduced several changes. An automatic slide-projector was used to present the pairs for the study-trials and the stimuli for test-trials. These were projected on a white wall a few feet in front of *S*. The presentation time was 3.75 sec. on both the study- and test-trials, with 2.5 sec. between successive slides. In the 30 sec. given between test- and study-trials and between study- and test-trials *S* worked on a simple cancellation-task.[13] As a final change the lists consisted of 16 pairs rather than the 12 pairs used in Experiment II A. Fifteen *S*s served in each of the three conditions. The *S*s were required to say aloud the stimulus- and response-terms on study-trials, and the stimuli on test-trials. If the original pool of items was exhausted for a given *S* in *DOC*, the order of replacement was determined by the order in which the items had been missed.

Results. Learning was terminated after 14 study- and test-trials or after one perfect recitation, whichever came first. Since all *S*s achieved at least 13 items correct on a single trial, the response-measure used was the number of trials required to reach 13 correct responses. The F (3.73) was significant beyond the 5% level, the means being 5.00, 6.87, and 4.60 for C_1 *DOC*, and C_2 respectively. The *t*s for C_1 vs. *DOC* and C_2 vs. *DOC* are significant beyond the 5% level, but the difference between C_1 and C_2 falls far short of significance. Essentially, therefore, the results of Experiment III are the same as those for Experiment II A.

EXPERIMENT IV

Although inter-trial rehearsal was controlled in Experiment III, the rate of presentation of successive items was still slow, a new pair being presented every 6.25 sec. In Experiments IV and V the rate of presentation was increased to levels which correspond to those normally employed in memory-drum studies.

---

[13] Underwood, Studies of distributed practice: VI. The influence of rest-interval activity, *J. exp. Psychol.*, 43, 1952, 329-340.

Procedure. A device employing a manually operated slide-changer was constructed whereby stimulus-cards could be presented at rapid rates. *S* viewed the items through a small aperture in the beaverboard-screen. The stimuli were simply typed on paper and inserted in slide-frames. A timer flashed a small light every 3 sec. as a signal for *E* to change the slide, this change being accomplished by a quick push on a lever. With this device the pairs were presented at a 3-sec. rate on study-trials and the stimuli at a 3-sec. rate on test-trials. The 30-sec. interval between trials was again filled by the cancellation-task. The three conditions were the same as in previous experiments, with 15 *S*s assigned to each. Each list consisted of 10 pairs. The *S* did not say the pairs aloud on study-trials nor the stimuli on test-trials.

Results. The results were essentially the same as for Experiments II A and III. The mean numbers of trials required to reach one perfect recitation were 6.87, 8.73, and 5.93 for C_1, *DOC,* and C_2, respectively. The *F* (4.68) is significant beyond the 5% level. Again the *t*s between the *DOC* and C_1 and C_2 are significant beyond the 5% level, but the two controls do not differ significantly.

EXPERIMENT V

Procedure. The same procedure was used as in Experiment IV except that the rate of presentation was 2 instead of 3 sec. Twenty *S*s served in each of the three conditions.

Results. Increasing the rate of presentation increased the number of trials to learn, but the relative differences between conditions did not change. The mean numbers of trials to learn were 11.90, 13.80, and 8.20 for C_1, *DOC,* and C_2, respectively. Again the *F* is significant beyond the 5% level. The *t* between C_1 and C_2 is 1.89, just short of the 5% level. The difference between C_2 and *DOC* is significant beyond the 1% level, while the difference between C_1 and *DOC* closely approaches the 5% level.

DISCUSSION

Experiment I, in which difficulty of response-learning was varied, showed clearly that the removal of missed items in *DOC* resulted in selection of easier items, although the procedure impeded over-all learning. Thus, *DOC* *S*s learned more slowly than C_1 *S*s, and C_1 Ss more slowly than C_2 *S*s. In Experiments II A through V, in which letter-number pairs were used and in which response-learning was minimized, the results were less clear-cut. A summary of these four experiments is shown in Table II. The entries are mean number of trials to reach a criterion of one perfect trial except in the case of Experiment III where the values refer to mean number of trials to reach 13 correct on a single trial.

The results for these four experiments are remarkably consistent; under a rather wide variety of procedures in the presentation of the same materials, the results show marked uniformity. In each of the four experiments the F fell between the 5%- and 1%-significance levels. In all four experiments the DOC resulted in slower learning than did the procedures for C_1 and C_2. The difference between C_1 and DOC achieved the 5% level twice, and twice it did not. In three of the four experiments the difference between DOC and C_2 exceeded the 1% level, and once the 5% level. Performance under C_2 was always better than under C_1 but in no experiment did this difference attain the 5% level of significance.

The consistency of the differences between conditions across all experiments would argue for the reliability of the differences. Furthermore, a 3 × 4 analysis of variance employing Experiments as one variable and Conditions as the other (with 80 *S*s in each condition), demonstrates the reliability of the over-all differences. The critical comparison shows that the *F*s for Conditions and Experiments exceed the 1% level; the F for interaction does not attain the 5% level. The fact that the interaction is not significant indicates that differences among conditions have not varied significantly for the four experiments. When the within-groups variance is used to estimate sigma of the difference, the t for the mean difference between C_1 and DOC is 2.46, and for the difference between C_1 and C_2 is 2.33.

TABLE II

MEAN NUMBER OF TRIALS TO CRITERION UNDER THREE CONDITIONS IN EXPERIMENTS II–V

Condition	Experiments				
	IIA	III	IV	V	Overall
Control 1 (C_1)	3.80	5.00	6.87	11.90	6.63
Drop-Out Condition (*DOC*)	4.40	6.87	8.73	13.80	8.03
Control 2 (C_2)	3.40	4.60	5.93	8.20	5.30

Previous investigators, failing to find a significant difference between C_1 and DOC, have concluded that associative learning between stimulus- and response-terms takes place in one trial. If associative learning takes place gradually, the argument has run, the DOC should require more trials to learn than is true for C_1. Since these investigators did not find a difference between C_1 and DOC, it was concluded that learning occurs on one trial. If such logic is accepted, the results of the present experiments would indicate that learning does not take place in one trial and that the gradual build-up of associative strength remains a tenable hypothesis. Such a conclusion, however, is no more defensible than the one originally drawn until full understanding of the implication of the 'drop-out method' is obtained.

Is there a fundamental conflict between our results and the results of those who have used the same materials and obtained 'no difference' between C_1 and *DOC?* Probably not. In two of our experiments (II A and V) the difference did not achieve acceptable levels of reliability and in Experiments III and IV the difference barely reached the 5% level of significance. More recent reports are quite in line with our findings.[14] Yet, if one wishes to assume a gradual buildup of associative strength, the question could still be raised as to why the learning under *DOC,* is not *seriously* impaired; taken at face-value, borderline statistical significance does not indicate serious impairment. The resolution of these problems lies, we believe, in item-selection and in the nature of learning under *DOC,* two issues to which we now turn.

Item-selection. We have shown that some item-selection occurred in Experiment II A in learning under *DOC.* The number of easy items appearing on the last trial of *DOC* is greater than the number present initially. Some further evaluation of this selection will be undertaken using the data from all four experiments in which letter-number pairs have been used.

One index of difficulty used for each pair is the ratio between the number of times an item appeared in the final *DOC* list and the number of times the item could have appeared. Product-moment correlations for item-difficulty of the 50 pairs as measured by this ratio for each item varied from 0.38 to 0.63 among the four experiments. All are significant beyond the 1% level. Clearly, the different experiments produce considerable communality among the items retained in learning under *DOC.*

As a second measure of item-difficulty we have used ratings of 'ease-of-learning.' A total of 29 *S*s, none of whom served as *S*s in the experiments, was given the 50 pairs printed in random order on a single sheet of paper. The *S*s were asked to rate the relative difficulty which they would experience in learning each item if all 50 pairs had to be learned. They were further instructed to read through the list once slowly before starting to make their ratings. The ratings were made along a 5-point scale with 1 representing the most difficult, 5 the least difficult.

These 'ease-of-learning' ratings have been amazingly accurate with other materials when tested in learning situations. With the present materials the relationships are not as high but still substantial. The correlations between the ratings and the selection-ratio described above varied from 0.53 to 0.62 for the four experiments. When a single ratio was obtained for each item by summing across all experiments, the product-moment correlation between these ratios and the ratings was 0.70. In subsequent analyses we will use the ratings as a measure of item-difficulty.

---

[14] G. R. Lockhead, A re-evaluation of evidence for one-trial associative learning, this JOURNAL, 74, 1961, 590-595; J. C. Reed and W. D. Riach, The role of repetition and set in paired-associate learning, this JOURNAL, 73, 1960, 608-611.

If there is selection of easier items under *DOC,* the mean rating should be higher for the C_2 lists than for the C_1 lists. A mean ratings was obtained for each list learned in C_1. Thus, each item in the list was assigned its appropriate rating, and the sum for all items in the list was determined. The sum of the ratings for all items was then divided by the number of items in the list to get a mean rating for that list. Across all experiments there were 80 C_1 lists, and 80 C_2 lists. The mean list-rating for C_1 was 2.98, for C_2, 3.25. The change is significant beyond the 1% level.

We will conclude that there is considerable agreement among *S*s on items which are easy and which are difficult, and that the procedures under *DOC* result in a final list which should be easier to learn than the list given initially. Clearly, therefore, there is a positive factor in learning under *DOC* which is not present in learning C_1.

Analysis of experimental procedure. Let us now consider the nature of the task presented *S* in learning under *DOC.* First, it will be remembered that the initial lists presented *DOC S*s are determined by drawing items randomly from a pool of 50 items. Sampling theory would lead us to expect, therefore, a distribution of list-difficulty. That is, by random sampling of items, we would expect some lists to contain more easy than difficult items, resulting in a list which should be relatively easy. Likewise, some lists formed by sampling should contain more difficult than easy items so that the list as a whole could be classed as difficult. Of course, the majority of the lists will be neither extremely easy nor extremely difficult since they will consist of about an equal number of easy and difficult items.

As the next step, consider a *DOC S* who is given an initially easy list. While he should learn several of the easy items on the first trial (especially when the rate of presentation is slow), his replacement-items on subsequent trials must necessarily be, on the average, more difficult than those he has already missed. Since the initial list consisted of more easy than difficult items, the remaining pool, from which replacements are drawn, must consist of more difficult than easy items. Hence, in drawing randomly from the pool, more difficult than easy items will be drawn. Beyond the first trial, therefore, learning should be impeded by the insertion of more difficult items and the *DOC S* may be expected to have more difficulty in mastering the list than the C_1 *S* in the same block who merely has to learn a list made up of easy items.

Next, consider a *DOC S* whose initial list is a difficult one. Performance on the first test-trial should be poorer than that of the *S* given an easy list initially. However, the *S* given the initially difficult list will, for replacement-items, be given more easy than difficult items so that learning should proceed rapidly following the first trial. Such an *S* may in fact learn by *DOC* more rapidly than the C_1 *S* in his block since the latter S *must* master a difficult list. Thus, *S*s having initially easy lists should show slow learning after the first trial, those having initially difficult lists should show rapid learning after the first trial; the result could be a cross-over of the acquisition-curves.

We shall now see how these expectations work out. For each *S* in the *DOC* an index of list-difficulty was obtained by using the ease-of-learning ratings. The method for determining this was explained earlier. From each of the four experiments the five *S*s given lists with the highest mean ratings (easy lists) and the five *S*s given lists with the lowest mean ratings (difficult lists), were selected. Across the four experiments this procedure produces 20 *S*s given lists having initially high mean ratings and 20 *S*s given lists with initially low mean ratings. The mean number of items correct on the first three learning trials was determined for these two groups independently. For the 20 *S*s given initially easy lists (high mean ratings) the mean numbers correct on Trials 1, 2, and 3 were 3.60, 4.70, and 5.60, respectively. For the 20 *S*s given the initially difficult lists the corresponding values were 2.75, 4.30, and 6.00. Clearly, the expected cross-over is occurring.

As a second test of the effect of drawing random lists, we have used all *S*s in all experiments. First, the distributions of mean list-ratings were divided at the median so that half the lists may be said to have high ratings (*H*-lists) and half low ratings (*L*-lists). In Experiment II A the division gave 15 *S*s having *H*-lists and 15 having *L*-lists. In Experiment III the division was 7 *H* and 8 *L;* in IV, 7 *H* and 8 *L*; and in Experiment V, 10 *H* and 10 *L*. Thus, 39 *S*s were given C_1 lists that are classed as easy, and 41 *S*s were given lists classified as difficult. Likewise, the initial lists for 39 *DOC S*s were easy and for 41 *S*s, difficult. It should be clear that the *S*s in C_1 and *DOC* are independent groups of *S*s, and that they were assigned randomly to conditions so that the four groups would be expected to be equivalent in learning ability. For each of these four groups we have two measures; namely, mean difficulty-ratings of the lists and mean number of trials to learn. In addition, we have determined the mean rating of the lists made up of items occurring on the last trial of learning by the *DOC S*s; these lists, of course, are the ones learned by C_2 *S*s.

The results are shown in the two graphs in Fig. 2. Fig. 2 A shows the change in mean ratings of the lists as a consequence of the process of selection occurring under *DOC*. Those *S*s starting with initially difficult lists show a much greater increase in mean ratings than do those starting with initially high ratings. The interaction is highly significant statistically as is also the over-all increase in mean ratings as a consequence of selection in the *DOC*.

Turning next to Fig. 2 B, another interaction is clearly present. Those *Ss* given initially difficult lists actually learn in slightly fewer trials under *DOC* than *Ss* who learn C_1. In the case of *Ss* given initially easy lists, however, over 3.00 trials more are required to learn under *DOC* than under C_1. The difference in the slope of the two curves (the interaction) is highly significant statistically.

There are a number of implications which follow from the above facts.

(1) Over-all, our results have shown learning under *DOC* to occur more slowly than under C_1. It is apparent from Fig. 2 B that this effect must be produced by only half the *Ss*, since the other half show a slight reverse effect.

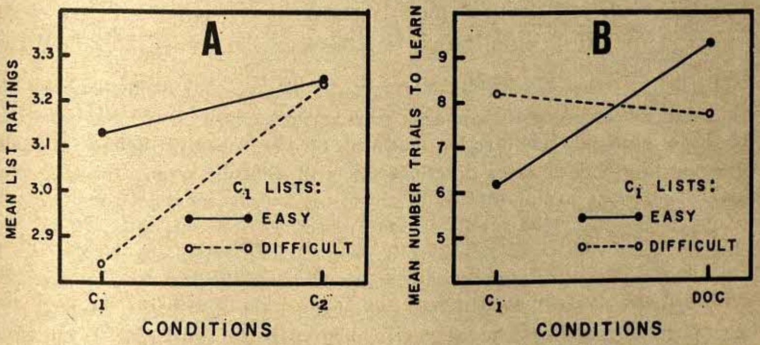


FIG. 2. RELATIONSHIPS BETWEEN C_1 AND C_2 RATINGS (A), AND C_1 AND *DOC* LEARNING (B) WHEN *SS* ARE DISTINGUISHED ON THE BASIS OF EASY AND DIFFICULT LISTS FOR C_1

(2) The C_2 lists in our experiments have been learned more rapidly than the C_1 lists, although the differences were smaller than were anticipated in view of the demonstrated selection of easy items. It is apparent that most of the selection occurs by *Ss* in *DOC* given initially difficult lists. The *Ss* that start with easy lists do not add much to this selection, as may be noted by the relatively small increase in mean list-ratings from C_1 to C_2. Thus again, it would appear that most of the difference in learning C_1 and C_2 is produced by only half the *Ss*. In actual fact, those *Ss* given the easy lists in C_1 required, on the average, only one more trial to learn than the C_2 *Ss* in corresponding blocks. The *Ss* given the difficult lists in C_1 required 2.07 more trials to learn than the C_2 *Ss* in corresponding blocks.

(3) It will be remembered that in Experiment II B three lists of eight pairs each were presented for learning on a memory-drum. One of these lists was used by Clark, Lansford, and Dallenbach.[15] These investigators, rather than drawing a

---

[15] Clark, Lansford, and Dallenbach, *op. cit.*, 35.

different random list for each *S* in C_1 and *DOC,* used a single random list for all *S*s. It was our belief that this single random list was a very difficult one, and our results showed that this list was a much more difficult list to learn than a list of eight of the most difficult items in Experiment II A. However, direct comparison of the ease of learning these lists is not completely justified; Clark, Lansford, and Dallenbach had paired their stimulus- and response-terms by a table of random numbers and their pairings were obviously not the same as ours. It is conceivable that the single list they used was representative of their pool, but in view of its demonstrated great difficulty and in view of other results they present, this seems highly doubtful. In any event, if this one list was constituted of several of the more difficult pairs in their pool, their results would be expected. They found that learning under *DOC* was about twice as fast as learning of the C_1 list. If this single list was a difficult one, the slope of the C_1 difficult-list curve in Fig. 2 B would be exaggerated, implying more rapid learning under *DOC* than under C_1. If the *S*s in *DOC* are given a very difficult list the process of selection will almost certainly introduce a number of easy items. The C_1 *S*s, on the other hand, have to learn the difficult list.

(4) The facts demonstrated in Fig. 2 make minor contradictions among the various experiments using letter-number pairs understandable. Each investigator has made his own pairings and each has drawn his own random lists. Different pairings may result in a pool of items differing widely in difficulty or in a pool of items fairly homogeneous in difficulty. If the items are homogeneous in difficulty, item-selection (other than idiosyncratic selection) becames a minor factor; if the pool is heterogeneous with regard to difficulty, item-selection becomes a major factor in determining the results. Or, if the distribution of item-difficulty is skewed in either direction the results will be affected. Furthermore, the distribution of mean list-difficulty actually drawn from the pool and given to C_1 and *DOC* *S*s will affect the results. Skewness of these distributions in one or the other direction will exaggerate or minimize the slope of the curves shown in Fig. 2 B. If the slopes of the two are equivalent, one upward and one downward, the over-all effect would be to produce no difference in trials to learn for *DOC* and C_1. We suspect that this has happened in the work of other investigators.

SUMMARY AND CONCLUSIONS

Several experiments were performed to evaluate item-selection in paired-associate learning. More particularly, the interest was centered around the procedures wherein items missed on a given trial are replaced by new items on the subsequent trial. Comparisons of learning under this condition were made with a condition in which the items remained constant from trial to trial. Previous investigators have found that learning is essentially equivalent under these two conditions. From his own studies showing such results, Rock has inferred that association between stimulus and response is an all-or-none affair; associative strength does not develop gradually. The reasoning has been that if associative strength develops gradually before the first correct anticipation occurs, learning under the drop-out condition

should take longer than under the control (constant items) since the removal of items which have some associative strength, but which have not been correctly anticipated, would retard learning.

One experiment in the present report was designed to show that when response-learning is a major factor, the 'drop-out procedure' will result in slower learning than the constant-list procedure. It was slower; furthermore, clear evidence was available for item-selection in the drop-out procedure since items given on the final trial of the drop-out condition were easier items, on the whole, than those given on the first trial.

Four additional experiments were performed with materials (letter-number pairs) used by several investigators and for which response-learning is minimal. An evaluation of the results from all of these experiments showed that the 'drop-out procedure' resulted in slower learning than the constant procedure. Item-selection was clearly present; the list of items appearing on the last trial in the drop-out procedure was an easier list than that given initially. Furthermore, it was shown that whether or not the drop-out procedure provided a more difficult task than the constant procedure depended upon the difficulty of the lists initially presented. If the list initially given consisted of a preponderance of easy items the drop-out procedure resulted in much slower learning than under the constant procedure because the replacement-items would be more likely to be difficult than easy. On the other hand, if *S* started with a list of preponderantly difficult items, replacements were more likely to be easy than difficult and learning may be as rapid or more rapid as compared with the constant condition.

Do our results deny one-trial learning? Of course not; but they strongly indicate that the method used to infer one-trial learning is quite inadequate for that purpose because of the item-selection that occurs in the drop-out procedure. One-trial learning could still be occurring but one cannot use these procedures to infer it. So also, gradual development of associative strength may be occurring. The situation is simply not one which allows a decision to be made between these notions.

REPETITION AND PAIRED-ASSOCIATE LEARNING

By Leo Postman, University of California, Berkeley

The role of repetition in the formation of verbal associations has recently become the focus of considerable discussion. The issue has been joined between two conceptions of the associative process—growth by successive increments vs. all-or-none change. The incremental theory asserts that each successive trial adds to the strength of an association, provided the conditions required for the reinforcement of a habit are met. If frequency of exposure is considered a sufficient condition of learning, the associations between stimuli and responses should be strengthened during repeated presentations of a list. By contrast, the all-or-none theory asserts that on any one trial associations are either fully formed or do not grow at all.[1] The apparent beneficial effects of repetition apply to the acquisition of a series but not to the formation of individual associations. Only a limited number of associations can be established on a given trial, and successive exposures to the series provide opportunities for increasing numbers of associations to be established, each on a single trial.

The question at issue is not whether a verbal association *can* be learned on a single trial. In experiments on rote-learning correct responses are frequently observed after one exposure to the list. It is consistent with an incremental theory that the rate of growth for some associations is sufficiently rapid to produce a correct response after one trial. The point under dispute is whether *all* associations are formed on a single trial. The disagreement comes to a head when one considers the effects of trials on which *S* fails to give the correct response. An incremental theory would have to hold that some associative growth occurs on such trials. According to an all-or-none theory, on the other hand, trials resulting in failure contribute nothing to associative growth.

Evidence allegedly supporting the all-or-none theory has been presented

---

* Received for publication January 23, 1961. This research was facilitated by a grant from the National Science Foundation.

[1] For a general discussion of the two positions see W. K. Estes, Learning theory and the new "mental chemistry," *Psychol. Rev.*, 67, 1960, 207-223. Estes identifies reinforcement of a habit with "the operation that is supplied by the experimenter in order to produce learning" (*ibid.*, 209). Under this definition, paired presentation of a stimulus and response constitutes a reinforcement, regardless of *S*'s response.

in a series of studies by Rock.[2] In the basic experiment on paired-associate learning, Rock compared a control group which learned a constant list of pairs with an experimental group for which all pairs missed on a given test-trial were replaced by new ones. This procedure insured that all pairs mastered by the experimental *S*s were learned in one trial whereas the control *S*s were exposed repeatedly to each pair. Using both letter-number pairs and pairs of nonsense-syllables, Rock found no differences between the two groups in the number of trials to a criterion of one perfect recitation of the list.[3] Rock concluded from these results that repetition does not contribute to the formation of associations.

Rock's findings, at least for the letter-number pairs, have been confirmed by Wogan and Waters, and by Clark, Lansford, and Dallenbach.[4] In discussing these results, both Rock and the other investigators recognized some important sources of bias inherent in the experimental design, and attempts were made to assess the effects of these biases in control-experiments.

A first critical question is whether Rock succeeded in controlling the frequency of repetitions under the conditions of his experiment. The validity of his conclusions hinges on the assumption that each of the pairs learned by the experimental *S*s had been repeated only once whereas most of the pairs learned by the control *S*s had been repeated more often. During the training trials each card was exposed for 3 sec., and there was a 5-sec. interval between successive cards. There was an interval of unspecified length between each training trial and test-trial, as well as a 30-sec. interval between each test-trial and subsequent learning trial. Thus, there was considerable opportunity for *S*s to rehearse the pairs. To the extent that there was such uncontrolled repetition, the basic conditions of the experiment were not met. Clark, Lansford and Dallenbach report that *S*s, in their experiment, used the intervals between exposures to construct and rehearse mnemonic devices; such rehearsal also took place between trials. These investigators also point out that the slow rate of presentation allowed an indeterminate number of repetitions.

To reduce the opportunities for uncontrolled repetition, Clark, Lansford, and Dallenbach performed an additional experiment in which the exposure-time for each pair was reduced to 1 sec., and the interval between successive pairs was eliminated. With a more rapid rate of presentation, the experimental *S*s learned the list significantly faster than did the control *S*s. Clark, Lansford, and Dallenbach conclude that uncontrolled rehearsal was not responsible for Rock's results and attribute the reversal to greater intra-list interference for the control group.[5] In a more recent study, Lockhead increased the rate of presentation to 0.75 sec. (with 1.25 sec. between cards) and reduced the intertrial interval to 12 sec. Under these conditions, he found a highly significant difference in favor of the control group, *i.e.* constant lists were learned faster than changing lists.[6] Regardless of rate, however, the question remains whether *S*s in both groups respond to each pair on each successive training trial. The *S*s are instructed to attend to all pairs presented during the training trials but there is no control over their responses. Thus, it would be quite possible for *S*s to concentrate on a few pairs at a time, rehearse them and essentially ignore other pairs. Such a procedure would be especially plausible for experimental *S*s who know that all pairs which they fail to learn will be removed from the list. This source of error can be reduced by achieving experimental control over *S*s' responses during the training trials.

---

[2] Irvin Rock, The role of repetition in associative learning, this JOURNAL, 70, 1957, 186-193; Irvin Rock and Walter Heimer, Further evidence of one-trial associative learning, this JOURNAL, 72, 1959, 1-16.

[3] Rock, *op. cit.*, 187-190.

[4] Michael Wogan and R. H. Waters, The role of repetition in learning, this JOURNAL, 72, 1959, 612-613; and L. L. Clark, T. G. Lansford, and K. M. Dallenbach, Repetition and associative learning, this JOURNAL, 73, 1960, 22-40. A more recent study (J. C. Reed and W. D. Riach, The role of repetition and set in paired-associate learning, this JOURNAL, 73, 1960, 608-611) shows that Rock's results may be peculiar to the nature of the instructions used in his experiment. The control group was told to associate stimuli and responses whereas the experimental group was instructed to learn all the pairs shown on a given trial. When the instructions were equalized, the experimental groups made more errors than the control group.

A second difficulty is that the experimental *S*s may learn a list of easier items than the control *S*s. The removal of all pairs which the experimental *S* fails to learn on a given trial certainly provides an opportunity for the selection of relatively easy items. Rock and Heimer report an experiment which was designed to determine whether the selection of items gives an advantage to the experimental group.[7] In the first phase of this experiment, the *S*s were given one exposure to 7 paired numbers followed by a matching test in which each stimulus was presented along with all the responses included in the list. In the second phase, a list of 12 paired numbers was presented 4 times, and a matching test was again used to measure learning. In the final phase of the experiment, only those *S*s were used who had learned no fewer than 2 but no more than 3 pairs of the short list, and no fewer than 3 but no more than 8 pairs of the long list. On a final critical trial these *S*s were exposed to a list of 8 pairs, 4 from the short list and 4 from the long list, all of which they had failed to learn previously. These two sets of 4 pairs were said to be of equal difficulty since each was drawn from the most difficult two-thirds of a prior list. In spite of the difference in the frequency of prior exposure, the two kinds of pairs were recalled equally well.

This procedure failed in its purpose of controlling the difficulty of the pairs for several reasons. (1) It is not reasonable to treat as equivalent the "most difficult two-thirds" of lists unequal in length and exposed for different numbers of trials. Rate of learning varies inversely with length of list.[8] One certainly cannot compensate for the length-difficulty relationship by a difference in the number of trials and then treat this same difference as the independent variable in assessing the effects of repetition on items said to be of equal difficulty. (2) A matching test on which both the stimuli and the correct responses are presented, the latter repeatedly, provides considerable opportunity both for rehearsal and for increases in intraserial interference. The magnitudes of these effects may have differed for the long and short lists. (3) The long list was learned after the short list. With highly similar materials this sequence of tasks probably produced associative and proactive inhibition of the second list as well as retroactive inhibition of the first list. The relative amounts of these interferences are not known. (4) "A considerable number" of *Ss* was dropped from the experiment because of failure to meet the criteria on the first two lists (2/7 and 3/12). Thus, if any conclusions can be drawn from this experiment, they would be limited to fast learners.

---

[5] Clark, Lansford, and Dallenbach, *op. cit.*, 35-38.

[6] G. R. Lockhead, A re-evaluation of evidence of one-trial associative learning, this JOURNAL, 74, 1961, 590-596.

[7] Rock and Heimer, *op. cit.*, 9-13.

[8] J. A. McGeoch and A. L. Irion, *The Psychology of Human Learning*, 1952, 489 ff.

Clark, Lansford, and Dallenbach used a different procedure to evaluate the possible bias introduced by the selection of items.[9] One week after serving in a replication of Rock's first study, both experimental and control *Ss* learned a second list by the conventional method (*i.e.* without replacement of failed items). The materials presented to each experimental *S* consisted of the first 12 pairs which he had failed to learn during the first part of the experiment and which had, therefore, been dropped from his list after a single trial. The control *Ss* learned a list selected at random from the pool of letter-number pairs.[10] Since the lists of the experimental *Ss* were selected for difficulty whereas the lists of the control *Ss* were chosen haphazardly, a difference in favor of the latter might be expected. However, the number of trials to criterion on the second list was identical for the two groups. Clark, Lansford and Dallenbach conclude that idiosyncratic differences in difficulty among the pairs cannot account for either Rock's or their results.

Both the experiments which we have discussed used failure to be learned on the first few trials as the criterion of the difficulty of a pair. A supplementary, and perhaps more direct, attack on the problem of item-selection is to ask whether the total list finally mastered by an experimental *S* is easier than that of a control *S*. A study by Underwood, Rehula, and Keppel using this design with lists of letter-number pairs presents clear-cut evidence for the selection of easy items by the experimental *Ss*.[11] The present study uses such a design with lists of nonsense-syllables.

Still another difficulty which complicates comparisons between the experimental and control groups in Rock's design concerns the conditions of intra-list interference. Intra-list errors, particularly misplaced responses, account for a substantial proportion of failures in paired-associate learning. Such errors may be quite persistent, especially when stimulus-similarity is high. The removal of all pairs which *S* does not learn in a single trial may systematically reduce the extent of intra-list interference. Rock and Heimer describe an experiment in which they attempted to check on this possibility.[12] Instead of removing the pairs which *S* failed to learn on a given trial, these investigators interchanged the response-terms of all such pairs. Although such a procedure is conducive to intra-list interference, the new experimental group did not differ from the control group in the original study. These results cannot be considered conclusive. The procedure used by Rock and Heimer does not control the conditions of rehearsal, nor does it eliminate the possibility that experimental *S*s selectively learn relatively easy pairs. Furthermore, the pattern of stimulus-response pairs changes from trial to trial, so that there is an opportunity for the selection of relatively easy *lists*. In contrast to the findings of Rock and Heimer, the results of Clark, Lansford, and Dallenbach indicate that intra-list interference is, indeed, greater for the control group than for the experimental group. Although there was no significant difference in the number of trials to criterion, the rate of errors was higher for control than experimental *S*s, as was the probability of failures following a correct response.[13]

---

[9] Clark, Lansford, and Dallenbach, *op. cit.*, 26-35.

[10] Immediately after completion of the first part of the experiment, that list had been presented to the control *Ss* and after a recall-test had been followed by a supplementary list of 12 new pairs. The purpose of this procedure was to equalize prior experience and retroactive interference for the two test-lists.

[11] B. J. Underwood, Robert Rehula, and Geoffrey Keppel, Item-selection in paired-associate learning, this JOURNAL, 75, 1962, 353-371.

[12] Rock and Heimer, *op. cit.*, 1-5.

Rock has called attention to the fact that there are other possible biases in his procedure which give an advantage to the control group.[14] First, the opportunities for familiarization with the stimuli and responses are greater with a constant than a changing list. The difference between the experimental and control groups was, however, no greater with nonsense-syllables than with familiar letters and numbers. Secondly, in most of the critical comparisons between the two groups a recall-test was used which required greater associative strength than, say, a matching test. An easier test might have revealed the presence of associations where the recall-test did not. Rock suggests that the experimental groups "were penalized by having such pairs eliminated." The point is not obvious since an easier test would presumably have improved the scores of both groups. Finally, missed pairs were removed from the list of an experimental *S* even if the failure was preceded by one or more correct responses. Thus, experimental *S*s were penalized for momentary oscillations whereas control *S*s were not. Rock states that such cases occurred only occasionally, and it should be possible to estimate the magnitude of the bias introduced by this procedure.

It is clear that the problems of control and interpretation in these experiments are exceedingly complex. Rock is disposed to argue that the biases operating for and against the experimental groups should balance each other. Even if they do not, "one might still be struck by the level of performance achieved by the experimental group without benefit of repetition."[15] It is difficult, however, to base any conclusions on either the absolute level of performance or the differences between the experimental and control groups as long as the magnitudes of the various biases are not known.

The purpose of the present study is to determine whether Rock's results obtained with lists of paired nonsense-syllables may reasonably be attributed to artifacts of procedure. When such materials are used, the mastery of the list requires the learning of unfamiliar responses as well as the association of these responses with the prescribed stimuli. Under these conditions the control group has a clear advantage since the integration of responses should be favored by repetition. Rock found that in spite of this bias the control group did not learn faster than the experimental group, and this finding was interpreted by him as strong evidence for one-trial learning. Actually, the lack of a difference is puzzling even if the one-trial hypothesis is assumed to be correct—since response-learning as well as association would appear to be independent of repetition. Thus, it becomes important to assess the influence of some of the variables which remained uncontrolled in Rock's experiment. Our special concern will be with *S*s' responses to the list during the training trials and with a further assessment of the extent of item-selection by the experimental *S*s.

---

[13] Clark, Lansford, and Dallenbach, *op. cit.*, 29-31.

[14] Rock, *op. cit.*, 190 f.

[15] *Ibid.*, 191.

METHOD

General. Two experiments were performed in which Rock's method of investigating one-trial learning was used. The major deviations from Rock's original procedure were (a) During the training trials the *S*s were required to spell both stimuli and responses aloud. Thus, *S*s were prevented from attending selectively to a few pairs at a time—a mode of attack which was possible under the conditions of the earlier studies. (b) Two Control Groups were included in the experimental design. Control Group I learned the lists presented initially to the Experimental *S*s and thus corresponded to the control groups in Rock's studies. Control Group II learned the lists finally mastered by the Experimental *S*s. A comparison between the two Control Groups permits a direct evaluation of the success of the Experimental *S*s in selecting relatively easy pairs, at least to the extent that the basis of selection was not idiosyncratic.

These modifications of Rock's design were used in both the present experiments. The procedure used on the test-trials differed in the two experiments. In Experiment I, the *S*s were reinforced on the test-trials, *i.e.* the correct pair was presented at the end of each anticipation-interval. This method of testing corresponds to that used in conventional studies of paired-associate learning. It has the advantage of counteracting the fixation of errors during the test-trials. The correction of errors may be especially beneficial to Experimental *S*s who are subject to cumulative associative interferences from discarded items. On the other hand, a second exposure to failed pairs may produce some learning which would be to the advantage of the Control *S*s but might interfere with the acquisition of new pairs by the Experimental *S*s. In view of these considerations no correction was given on the test-trials of Experiment II.[10] A comparison between the two experiments will make it possible to assess the effects of correction.

---

[10] It is not clear whether correction was given on the test-trials in Rock's first experiments. No reference to a correction is made in the original report. However, in describing a study which is said to have followed the original procedure in all respects, Rock and Heimer state that *S* was told whether he was right or wrong (*op. cit.*, 3).

Materials. Each *S* learned a list of eight paired nonsense-syllables. Syllables with association-values of 40.0%, 46.7%, 53.3% and 60.0% were used.[17] A total of 202 pairs was constructed. Stimuli and responses were drawn at random from the pool of syllables, with the restriction that the two members of a pair have no letters in common. The syllables were typed in capitals on white cards 3 × 5 in. in size. The stimulus and response appeared together on one side of the card, and the stimulus only on the other side.

Experimental design. The design for both experiments comprised 18 blocks of 3 *Ss*—one member each of the Experimental Group and Control Groups I and II per block. A different initial list, drawn at random from the pool of pairs, was used for each Experimental *S*. Pairs failed on a test-trial were replaced by new ones, also drawn at random, on the next training trial. Within each block, the member of Control Group I learned the initial list of the Experimental *S*, and the member of Control Group II learned the list on which the Experimental *S* reached the criterion of one perfect recitation. If the Experimental *S* failed to reach criterion, the list yielding the highest number of correct responses was used. If the highest score occurred on more than one trial, the earliest of these trials was used for selection of the list. The list of pairs remained constant, of course, for both Control Groups.

The member of Control Group II had to be run after the Experimental *S*. Thus, there were three possible orders of assignment to conditions within a block. The 18 blocks of the experiments were subdivided into 6 sub-groups of 3 blocks each. Within a sub-group of blocks, each of the orders of assignment was used once, the sequence of orders being determined by a table of random numbers. Thus, each sub-group of 3 blocks constituted a complete replication of the experiment. All *Ss* were undergraduate students who did not know the purpose of the experiment.

Procedure. The instructions described the procedures of training and testing and also included a statement that some of the pairs might be replaced in the course of the experiment. The same instructions were given to all *Ss*, so that the groups did not differ in their initial set.

During the training trials each pair was presented for 6 sec., during which *Ss* spelled both the stimulus and the response letter by letter. A longer exposure was used than in the earlier studies in order to allow *Ss* adequate time to spell both members. The *Ss* readily adopted the appropriate rhythm so that each period of exposure was entirely filled by the spelling of the two syllables. The interval between successive pairs was 3 sec. There was an interval of 45 sec. between training trials and test-trials, during which the stimulus-cards were shuffled and the new order of presentation recorded by *E*. There was also a 45-sec. interval between a test-trial and subsequent training trial during which the cards were again shuffled and the necessary substitutions were made in the case of the Experimental *Ss*.

In both experiments a 3-sec. interval of anticipation was used on the test-trials, during which the stimulus-member of the pair was presented to *S*. In Experiment I each anticipation-interval was followed by a 3-sec. presentation of the correct pair, and 3 sec. thereafter the next test-stimulus was presented. In Experiment II no correction was given, and a new stimulus was presented 3 sec. after the end of the anticipation-interval.

---

[17] The number of stimuli and responses of each association-value varied from *S* to *S* but tended to be proportional to the frequencies of these values in Glaze's norms. See J. A. Glaze, The association value of nonsense syllables, *J. genet. Psychol.*, 35, 1928, 263-264.

Practice was continued to a criterion of one perfect recitation or for 12 trials, whichever took the longer. If *S* failed to reach criterion on the 16th test-trial, the experiment was discontinued. Thus, data from all *S*s are available for 12 trials, regardless of whether and how fast they reached criterion.

RESULTS

(1) Learning to criterion. In Experiment I, 10 *S*s in the Experimental Group, 17 *S*s in Control Group I and all 18 *S*s in Control Group II reached the criterion of one perfect recitation. The proportion of *S*s reaching criterion is significantly lower in the Experimental Group than in the combined Control Groups ($\chi^2 = 12.15$, $df = 1$, $p < 0.01$).

TABLE I

SUMMARY OF LEARNING SCORES

Condition	Trials to criterion		Correct on Trials 1–12		Length integrative stage*	Length associative stage†
	Mdn	*Q*	*M*	*SD*		
Experiment I:						
Experimental	11.5	—‡	42.2	25.7	—	—
Control I	9.0	3.5	58.6	18.0	4.2	0.5
Control II	7.0	3.0	68.1	13.7	3.6	0.2
Experiment II:						
Experimental	>15.0	—‡	30.7	21.1	—	—
Control I	8.5	3.2	58.6	14.3	4.2	0.3
Control II	10.0	—‡	56.4	16.1	4.4	0.3

* Mean trial on which correct responses were first given anywhere in the list.
† Mean number of trials separating first occurrence of correct response anywhere and first correct anticipation.
‡ Indeterminate since 4 or more *S*s failed to reach criterion.

The median numbers of trials to criterion are shown in Table I. The rate of learning was slowest for the Experimental Group, and Control Group II surpassed Control Group I. Friedman's ranked analysis of variance shows the variation among the groups to be highly significant ($\chi^2 = 14.04$, $df = 2$, $p < 0.01$).[18] When the sign-test is used for individual comparisons, each group is found to differ significantly from the other two ($p = 0.05$ or less).

In Experiment II, 9 *S*s in the Experimental Group, 15 *S*s in Control Group I and 14 *S*s in Control Group II reached criterion. The difference between the proportions of *S*s in the Experimental Group and the combined Control Groups is again significant ($\chi^2 = 4.01$, $df = 1$, $0.02 < p < 0.05$).

---

[18] Frank Wilcoxon, *Some Rapid Approximate Statistical Procedures*, 1949, 7 f.

The median numbers of trials to criterion in Experiment II are presented in Table I. The median score of the Experimental Group could not be determined exactly since half the *Ss* failed to reach criterion; it is, therefore, listed as >15 (the score of the slowest *S* to reach criterion was 15).

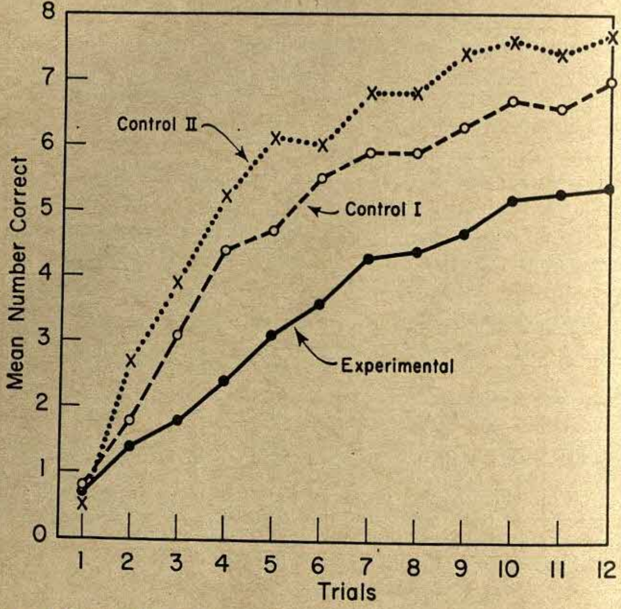


FIG. 1. MEAN NUMBERS OF CORRECT RESPONSES ON TRIALS 1-12 IN EXPERIMENT I

The Experimental Group again learned much more slowly than the Control Groups. However, the difference between the two Control Groups found in Experiment I is no longer present. The median of Control Group II is, in fact, slightly higher than that of Control Group I. The variations among groups is again significant by ranked analysis of variance ($\chi^2 = 6.24$, $df = 2$, $0.02 < p < 0.05$).[19] Sign-tests show the Experimental Group to differ significantly from both Control Groups ($p = 0.05$), whereas the difference between the latter is not significant.

(2) Learning curves. Fig. 1 shows the mean numbers of correct re-

---

[19] In a few cases, two *Ss* in the same block failed to reach criterion. Ranks were then assigned on the basis of the highest criterion reached in learning. In the two cases involving an Experimental and Control *S* these criteria were 1 and 6, and 2 and 7, respectively.

sponses on the first 12 test-trials in Experiment I. The differences on Trial 1 are not significant. The three groups are clearly separated thereafter. The Experimental Group progressed most slowly and Control Group II fastest, with Control Group I occupying an intermediate position. The mean numbers of correct anticipations given during the 12 trials are shown in Table I. The Experimental Group not only gave fewer correct responses, but also was considerably more variable, than the Control Groups. Since there was significant heterogeneity of variance, non-parametric tests were used to evaluate the differences in the numbers of correct anticipations. The variation among the groups is highly significant, as tested by a ranked analysis of variance ($\chi^2 = 12.36$, $df = 2$, $p < 0.01$). According to Wilcoxon's test for signed ranks, each group differs significantly from the other two at the 5% level or better.

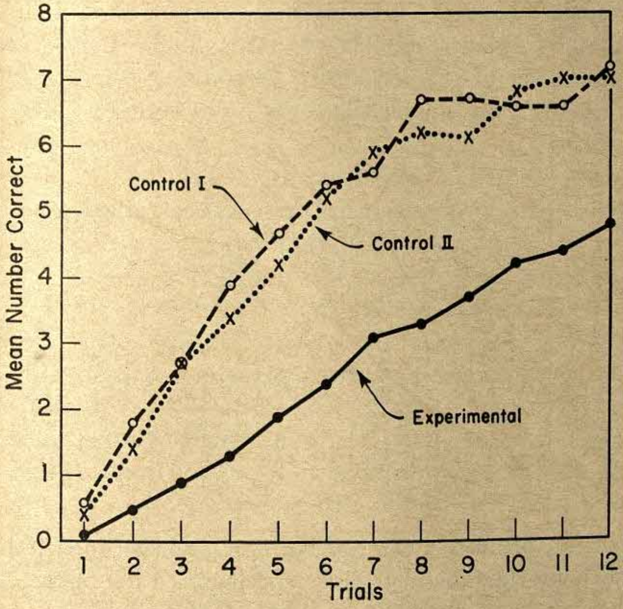


FIG. 2. MEAN NUMBERS OF CORRECT RESPONSES ON TRIALS 1-12 IN EXPERIMENT II

The learning curves for Experiment II are shown in Fig. 2. The differences on Trial 1 are again not significant. As in Experiment I, the rate of improvement is clearly slowest for the Experimental Group. In contrast to the earlier results, however, the curves of the two Control Groups overlap closely. The mean numbers of correct anticipations are presented in

Table I. Ranked analysis of variance again shows the variation among groups to be significant ($\chi^2 = 12.12$, $df = 2$, $p < 0.01$). The Experimental Group differs significantly from both Control Groups ($p < 0.01$ as determined by tests for signed ranks). There is no reliable difference between the two Control Groups.

Attention has been called to the fact that Experimental *S*s are penalized by the procedure of removing failed pairs which had been given correctly on earlier trials. To correct for this bias in the learning scores, an additional analysis of the data on acquisition was made. Table II shows the mean cumulative number of *different* correct responses given on Trials 1-12. In this tabulation, *S*s are given credit for each new correct response, regardless of whether or not it was given correctly again on subsequent trials. This analysis provides the most direct and unbiased test of the effects of repetition since it reflects the rate at which *new* associations are added under the various conditions. It is clear that the picture is the same as that obtained by the conventional method of scoring—the Experimental Group consistently falls below the Control Groups.

TABLE II

MEAN CUMULATIVE NUMBER OF DIFFERENT CORRECT RESPONSES ON TRIALS 1–12

Condition	Trial											
	1	2	3	4	5	6	7	8	9	10	11	12
Experiment I:												
Experimental	0.7	1.7	2.2	3.0	3.7	4.4	5.1	5.5	5.9	6.4	6.6	6.8
Control I	0.6	1.9	3.5	4.9	5.6	6.1	6.6	6.8	7.2	7.4	7.6	7.6
Control II	0.5	2.7	4.3	5.5	6.6	7.1	7.5	7.7	7.7	7.8	7.8	7.9
Experiment II:												
Experimental	0.2	0.7	1.4	1.7	2.4	3.2	4.1	4.7	5.2	5.7	6.0	6.4
Control I	0.6	1.8	2.9	4.3	5.5	6.3	6.9	7.1	7.2	7.3	7.3	7.4
Control II	0.4	1.4	2.9	4.2	5.1	5.9	6.7	7.1	7.3	7.6	7.6	7.8

All the analyses agree in showing that in both experiments the Experimental Group learned much more slowly than the Control Groups. The two Control Groups differed only in Experiment I. Comparison between the two experiments indicates that the change from correction to noncorrection during the test-trials was more detrimental to the Experimental Group than to Control Group I. The increase in the difference between the Experimental Group and Control Group I was accompanied by the loss of advantage for Control Group II.

(3) Stage-analysis. Since a significant difference between the two Control Groups was found in Experiment I, the question arises as to the

specific characteristics of acquisition which are responsible for this difference. A stage-analysis was, therefore, performed on the data of the Control Groups in both experiments. In such an analysis, a distinction is made between two stages of acquisition, viz. (a) the integrative stage, and (b) the associative stage.[20] During the integrative stage the prescribed units are integrated and become available as responses to *S*. The duration of this stage is estimated by the number of the trial on which a response is first given regardless of whether or not it is paired with the correct stimulus. The duration of the associative stage is estimated by the number of trials that elapse between the time the response is first given and the time it is paired with the correct stimulus. The two stages undoubtedly overlap, and there is good reason to believe that the integrative stage continues beyond the point of first correct association. However, the two indices make it possible to assess the relative importance of the two stages for different materials and conditions. Table I shows that in Experiment I there is a difference in favor of Control Group II in both stages of learning. However, by Wilcoxon's test for signed ranks only the difference in the duration of the associative stage is significant ($p < .05$). Thus, there appears to have been selection of pairs for ease of association, and probably for availability of responses. In Experiment II, the two Control Groups are all but indistinguishable by either measure.

(4) Distribution of errors. Incorrect responses were classified into failures to respond and overt errors. The latter were in turn divided into (a) misplaced responses, (b) stimulus-intrusions, (c) mixed errors, and (d) partial responses. A misplaced response occurred when a correct syllable was given to an inappropriate stimulus. An error was classified as a stimulus-intrusion when one of the stimuli in the list was given as a response. Mixed errors include all three-letter responses which were not items in the list. Partial responses consisted of either one or two letters.

For each *S*, the number of overt errors was expressed as a percentage of the total number of incorrect responses. In Experiment I, the mean percentage of overt errors was 49.0 for the Experimental Group, 41.3 for Control Group I, and 42.4 for Control Group II. Thus, the relative frequencies of overt errors were comparable for the three groups. There were, however, substantial differences among the groups in the distribution of overt errors. Table III shows the mean percentages of the various classes of overt errors. The percentage of misplaced responses is highest for Control Group I and lowest for the Experimental Group. Control Group II occupies an intermediate position but is closer to the Experimental Group than to the other Control Group. The pattern is the same for stimulus-intrusions although such errors occurred rarely. Mixed errors and partial errors occurred with relatively higher frequency for the Experimental Group and Control Group II than for Control Group I.

---

[20] See B. J. Underwood and R. W. Schulz, *Meaningfulness and Verbal Learning*, 1960, 92-94.

The variation in the percentage of misplaced responses is significant, as tested by ranked analysis of variance ($\chi^2 = 11.36$, $df = 2$, $p < 0.01$). Tests for signed ranks show that Control Group I differs from both the other groups at the 1% level. The difference between the Experimental Group and Control Group II is not significant.

TABLE III

MEAN PERCENTAGES OF DIFFERENT TYPES OF OVERT ERRORS

Condition	Misplaced responses	Stimulus-intrusions	Mixed errors	Partial responses
Experiment I:				
Experimental	19.7	3.0	56.7	20.6
Control I	39.9	8.9	38.6	12.6
Control II	26.7	3.8	52.6	16.9
Experiment II:				
Experimental	13.9	4.6	72.3	9.2
Control I	31.8	2.1	55.5	10.6
Control II	30.5	5.3	51.5	12.6

In Experiment II, the relative frequency of overt errors was somewhat lower than in Experiment I but again comparable for the three groups. The mean percentage of overt errors was 38.7 for the Experimental Group, 36.7 for Control Group I, and 38.0 for Control Group II. The distributions of overt errors are presented in Table III. The percentage of misplaced responses is considerably lower for the Experimental Group than for both Control Groups. For mixed errors, the differences are in the opposite direction. The distributions of errors for the two Control Groups are quite similar. The significance of the differences in the percentages of misplaced responses was tested as before. The over-all variation is significant ($\chi^2 = 8.36$, $df = 2$, $0.01 < p < 0.02$), and the Experimental Group differs significantly from Control Group I ($p < 0.01$) and Control Group II ($p < 0.05$).

(5) Evidence for selection of items. In the present design, the extent to which the Experimental *Ss* succeeded in selecting relatively easy lists is gauged by a comparison of the two Control Groups. A difference in favor of Control Group II points to a systematic selection of items or lists.

According to this criterion, there was systematic selection of items in Experiment I but not in Experiment II. Two problems now require attention: (a) the probable bases of selection where it did occur, and (b) the reason for the difference between the two experiments.

(a) Characteristics of selected items. A comparison of the lists learned by the two Control Groups may provide some information about the nature of the items selected by the Experimental *S*s. Such an analysis is limited by the fact that about half the Experimental *S*s failed to reach criterion. In such cases whatever selection occurred was only partial. The findings concerning the characteristics of selected items remain essentially the same, however, whether the entire Control Groups are considered or only those blocks in which the Experimental *S* reached criterion. Hence the tabulations to be presented below include the data for all *S*s. It must be emphasized that the comparisons between the lists of the two Control Groups are made *ex post facto,* and any conclusions drawn from the analysis are therefore, highly tentative.

TABLE IV

DISTRIBUTION OF DIFFERENCES BETWEEN LISTS OF CONTROL GROUPS I AND II

(Entries are numbers of cases)

Variable	Experiment I			Experiment II		
	CII>CI	CII=CI	CII<CI	CII>CI	CII=CI	CII<CI
Mean association-value of stimuli	11	—	7	13	—	5
Mean association-value of responses	15	—	3	8	1	9
Number different letters in stimuli	6	5	7	7	4	7
Number different letters in responses	10	2	6	7	—	11

The lists learned by the two Control Groups were compared on two task-variables, viz. (i) mean association-value and (ii) intra-list similarity. Each characteristic was evaluated separately for stimuli and responses. Table IV shows the distributions of differences between the two Control Groups on each characteristic, *i.e.* the frequency with which Control Group II was superior, equal, and inferior, to Control Group I.

The opportunity for selection of items on the basis of association-value was limited since the total range of variation was between 40% and 60%. The distributions of differences in Table IV suggest that there was some selection of stimuli for association-value, especially in Experiment II. Selection of responses for association-value appears to have occurred in Experiment I but not in Experiment II. It is known that learning is influenced much more by meaningfulness of responses than of stimuli. Hence a larger difference in speed of learning between the two Control Groups would be expected in Experiment I than in Experiment II. The results are consistent with this expectation.

For both stimuli and responses, the number of different letters making up the eight items was used as the index of intra-list similarity. There is little evidence for systematic reduction in intra-list similarity, with the possible exception of response-similarity in Experiment I. In the learning of nonsense-syllables, similarity of responses has less pronounced effects than similarity of stimuli, and it is doubtful that it was an important basis of selection in Experiment I.

The one hypothesis which is clearly suggested by the comparison of the initial and terminal lists is the selection of pairs for meaningfulness of responses. The results of the stage-analysis are consistent with this hypothesis.

(b) Conditions of selection. There is some evidence for selection of items in Experiment I but not in Experiment II. This discrepancy between the two experiments, which had not been expected, cannot be attributed to the difference between the two Experimental Groups in speed of learning. The mean criteria finally reached, and hence the numbers of correct items used in the construction of the lists for Control Group II, were comparable in the two experiments—6.1 in Experiment I and 6.0 in Experiment II. The selection of items occurred later in Experiment II, but the opportunities for detecting the presence of selected items were equal in the two cases.

One possible reason for the apparent failure of selection in Experiment II may be a higher level of cumulative inter-trial interference in the absence of correction. For purposes of estimating the relative amounts of inter-trial interference for the two Experimental groups, a new analysis of errors was made. We considered all misplaced responses and stimulus-intrusions on the trials up to the terminal criterion, including errors from inside the list and from prior lists. The percentage of such errors coming from prior lists was 22.8 in Experiment I, and 36.1 in Experiment II. The contribution of inter-trial interference to other overt errors and failures to respond cannot be assessed, but these percentages suggest that it may well have been greater in Experiment II. Thus, the acquisition of new items in Experiment II may have been determined more by the conditions of inter-trial interference than by the intrinsic difficulty of the items themselves. Selection of items on this basis would, of course, not be reflected in the performance of Control Group II.

There may, of course, be other reasons for the discrepancy between the two experiments, nor can the possibility of a chance difference be discarded. The question can be settled only by manipulating the opportunities for selection systematically under different conditions of learning.

Discussion

Under the conditions of the present study the Experimental Group learned significantly more slowly than the Control Groups. Thus, the results are clearly contrary to those of Rock. If the earlier results are interpreted as positive evidence for the hypothesis of one-trial learning, that interpretation is called in question by our findings.

The major difference between Rock's procedure and ours lies in the control over *S*s' responses during training. In Rock's experiment *S*s merely observed the pairs; in the present study they were required to spell both stimuli and responses. The requirement that *S* respond overtly to each pair during training insures that the effective length of the list is the same for all *S*s in both the Experimental and Control Groups. Without such a requirement, *S* can vary the number of pairs which he reads and rehearses at will from trial to trial. It is reasonable to assume that under Rock's conditions such selective rehearsal will occur more often in the Experimental Group than in the Control Group, especially when the learning materials are unfamiliar and difficult. Since time-per-item increases with length of list, the Experimental Group would have an advantage. The requirement of overt responses to each pair also reduces the opportunities for uncontrolled repetition of those pairs *S* is attempting to learn. In the present experiments such rehearsal remained possible, however, during the intervals separating training trials and test-trials.

The probable importance of uncontrolled rehearsal is indicated by a comparison between the corresponding groups in Rock's study using nonsense-syllables and the present Experiment II (in which the procedure was essentially similar to Rock's). In Rock's experiment the median number of trials to criterion was the same for both groups, viz. 8.5. In our Experiment II, the median of Control Group I was identical with that obtained by Rock, *i.e.* 8.5. On the other hand, the median of the Experimental Group exceeded 15 trials, with half the *S*s failing to reach criterion. It is fair to conclude that our *S*s in Control Group I behaved very much like Rock's, whereas our Experimental *S*s clearly did not.

The analysis of errors supports the conclusion that the course of learning was essentially different for the Experimental and Control Groups. The Control *S*s consistently gave a higher percentage of misplaced responses than did the Experimental *S*s. The relative frequency of such errors is known to vary directly as a function of familiarization with the responses in a list of paired associates.[21] It is likely that the differences in the percentages of misplaced responses reflect a higher degree of response-learning by the Control *S*s. In Experiment I, Control Group II gave relatively fewer misplaced responses than Control Group I. This finding points to selection of items for ease of associative connection. The results of the stage-analysis support these conclusions.

A significant difference in favor of the Control Groups is obtained whether or not correction is used on the test-trials. It had appeared possible that correction on the test-trials might bias the results against the Experimental Group since a second exposure to failed items could interfere with the acquisition of new pairs. The results suggest that correction reduced the fixation of errors and hence the amount of interference from prior discarded items. Thus, the net effect of correction seems to have been beneficial to the Experimental Group. The critical fact is, of course, that the differences between the Experimental Group and the Control Groups are significant in both experiments. Moreover, the size of the difference between the Experimental Group and Control Group II, which corrects for whatever normative selection of items does occur, is almost identical in the two cases. As measured by the mean numbers of correct responses on Trials 1-12 (Table I), this difference is 26.1 in Experiment I, and 26.3 in Experiment II.

---

[21] Underwood and Schulz, *op. cit.*, 100-126.

The findings concerning the systematic selection of items by the Experimental *S*s are ambiguous. Evidence for such selection, possibly based on meaningfulness of the responses, was obtained in Experiment I but not in Experiment II. It is obvious that the specific conditions under which the Experimental *S*s are trained and tested will influence the amount and kind of selection that will occur. Analysis of the errors suggests that without correction cumulative interference from discarded prior items may become decisive in the selection of items, at the expense of relatively small variations in meaningfulness. Whether or not selection occurs, it fails to overcome the significant advantage of Control Group I over the Experimental Group. The same conclusion applies to possible selections based on idiosyncratic factors which could not be assessed in this study.

The results of the experiments throw serious doubt on the validity of Rock's evidence for one-trial learning, at least as far as his findings with nonsense-syllables are concerned. Our results cannot be interpreted, however, as evidence against the hypothesis or in favor of incremental theory. There is every reason to believe that the superiority of the Control Groups is at least in part a function of the degree of response-learning. With repeated presentations, responses become increasingly well integrated and available for association. We referred earlier to the distinction between the integrative and associative stages of learning. Rock's procedure gives an inevitable advantage to the Control Groups in the integrative stage. The one-trial hypothesis is, however, concerned exclusively with the associative stage. Thus, any experimental comparison which can be influenced by differences in response-learning does not permit a clear-cut evaluation of the hypothesis. When letter-number pairs are used, response-integration may no longer be a critical factor, but with such materials there is now conclusive evidence for systematic selection of items.[22]

---

[22] Underwood, Rehula, and Keppel, *op. cit.*, this JOURNAL, 75, 1962, 353-371.

We conclude that the studies using Rock's design, including the present one, have yielded results which do not permit a decision between the one-trial hypothesis and the incremental hypothesis. A crucial experiment pitting the incremental hypothesis against the one-trial hypothesis remains to be designed.[23]

SUMMARY

Two experiments are reported investigating the role of repetition in the learning of paired nonsense-syllables. A procedure introduced by Rock was used: a control group learns a constant series of pairs, whereas for an experimental group all pairs failed on a given trial are removed and replaced by new ones. Two departures from the method used in the earlier study were introduced: (a) The *S*s were required to spell the items during training whereas they had merely observed them in the previous experiment. (b) Two Control Groups were used, one of which learned the initial lists of the Experimental *S*s, whereas the other learned the terminal lists. The second Control Group was used to assess the amount of systematic selection of easy items. In Experiment I correction was given on the test-trials. In Experiment II no correction was given.

In both experiments the Experimental Group learned significantly more slowly than the Control Groups. In contrast to Rock's results, repetition was found to have a significant effect on rate of learning. These findings underscore the importance of controlling *S*s' responses to the materials during the training trials. Evidence for systematic selection of items was obtained in Experiment I but not in Experiment II. Comparison of the two experiments suggests that the amount and kind of item-selection varies with the conditions of inter-trial interference.

Analysis of *S*s' performance indicates that the superiority of the Control Groups must be attributed, at least in part, to the beneficial effects of repetition on the integration of responses. The effects of repetition on the associative stage of learning remain indeterminate. Serious doubt is thrown on the validity of Rock's evidence for one-trial learning. The adequacy of his design for testing the hypothesis of one-trial learning is called in question.

---

[23] Recently Estes (*op. cit.*) has presented some challenging arguments and experimental data which he interprets as evidence for the one-trial conception. A discussion of this work is beyond the scope of this paper.

TEMPORAL AND SPATIAL EFFECTS OF SERIAL POSITION

By ARTHUR R. JENSEN, University of California, Berkeley

With few exceptions, research and theory concerning serial rote-learning have developed within the *S–R* connectionist tradition of Ebbinghaus, Thorndike, and Pavlov.[1] Thus, the principal explanations of serial phenomena involve the concepts of *S–R* bonds, remote associations, stimulus-traces, trace-conditioned responses, inhibition of delay, gradient of reinforcement, stimulus-generalization, intralist interference, response-competition, proactive and retroactive inhibition, etc. The various *S–R* theories as explicitly formulated all have in common the assumption that serial learning consists essentially of the acquisition of connections between items in the series, each item in turn serving as the stimulus for the anticipation of the next item. The picture is that of a chain of *S–R* connections.

According to this view, the explanation of certain serial phenomena, such as the skewed bowing of the serial-position curve, is to be sought in the nature of the *S–R* connections, either in terms of their relative strengths as affected by the interaction of excitatory and inhibitory processes that are assumed to build up in the course of learning, or by interference or response-competition between the items resulting from the formation of remote associations or from stimulus-generalization. Generally regarded by theories based on a paradigm of conditioning as one of the variables responsible for the bowed shape of the serial-position curve is the *temporal aspect* of the serial list, controlled by presenting the individual items serially in the aperture of a memory-drum at a given rate.

Purpose. The following experiments were addressed to the question of whether the occurrence of the serial-position effect depends upon a temporal serial presentation of the items or if the bowing is a more general phenomenon which may also occur when the ordering of the items is predominantly spatial rather than temporal and when *S*'s response is not limited to serial anticipation. To answer this question, four experiments were performed. In Experiment I, nine stimuli in a row were presented simultaneously; the series was predominantly *spatial*. This arrangement

---

* Received for publication December 9, 1960.

[1] A review of theories of the serial-position effect is found in J. A. McGeoch and A. L. Irion, *The Psychology of Human Learning*, 1952, 125-134.

admittedly would not rule out the possibility of temporal effects resulting from *S*'s visually scanning the stimuli in a certain order. In Experiment II there was a one-at-a-time presentation of the nine stimuli, always in one location, but there was *no* serial order. The stimuli appeared in a *random* order, except for some restrictions made for methodological reasons to be explained later. The spatial serial arrangement was in the *S*'s *response*-alternatives, which consisted of a row of nine buttons. The *S* had to learn by trail-and-error which button was associated with the particular stimulus presented at random on the screen in front of *S*. The reinforcement for pressing the correct button was the sound of a 'bong.' With this procedure, the idea of serial order probably never occurred to most of the *S*'s. In Experiments III and IV, intended as controls, the stimuli were presented in the usual temporal serial manner.

EXPERIMENT I

Subjects. All *S*s in these four experiments were upper-division and graduate students in the Department of Education. There were 30 *S*s in Experiment I. They were naïve with respect to psychological theory and experimentation. The proportion of women to men was about two to one.

Apparatus. The stimulus-items consisted of nine brightly colored simple geometric forms made of plastic.[2] The objects are ¼ in. in thickness and approximately 1 in. in height or diameter; they appear subjectively of about equal size. The nine objects were: red triangle, green square, blue triangle, yellow diamond, red circle, yellow triangle, blue circle, pink square, and green diamond.

Procedure. *S* was first shown all the objects in a randomly scattered arrangement and was told that the objects would be arranged in a row in front of him. He would be allowed to study the order for a period of 10 sec., after which the objects would be disarranged, and *S* would be required to try to arrange them in the proper order.

Flat on the table before *S* was taped a sheet of heavy white paper, 12 × 18 in. Centered on this sheet of paper was a 1½ × 12 in. rectangle, drawn in a heavy black line; the rectangle was in a horizontal position with respect to *S*. After *E* had displayed the objects to *S* and explained the task, *E* placed a black cardboard-screen in front of *S* so as to cut off *S*'s view while the objects were arranged in a particular order within the rectangle on the sheet of paper. The objects were evenly spaced about ½ in. apart. When *E* said "Ready" and removed the screen, *S* studied the series for 10 sec. *E* then placed the screen in front of *S* and completely disarranged the objects in a haphazard manner. Then the screen was removed and *S* was told to reconstruct the order of the objects. There was no time limit, and *S* was required to guess when in doubt. If *S*'s reconstruction was not perfect (it rarely was on the first trial), the entire procedure was repeated, until *S* attained the criterion of perfect reconstruction of the serial order. After each trial, while the screen was in front of *S*, *E* recorded *S*'s performance. Never were *S*'s mistakes pointed out to him; he was merely told that he would have to repeat the task until he got it perfect.

---

[2] These objects are on the market as "Child Guidance Toys, No. 709, Primary Fit n' Form, Ages 1-3 yrs.," manufactured by Archer Plastics, Inc., Bronx 72, N.Y.

A different order of items was used for every *S* that effects arising from any particular order would tend to be randomized in the over-all results. A restriction was placed on the formation of the series: no two objects of the same shape or of the same color could be adjacent to each other.

Results. The *S*s were able to learn this task quite easily. The mean number of trials to criterion was only 2.96; the mean number of errors was 9.06; and the mean percentage of errors was 33.96.

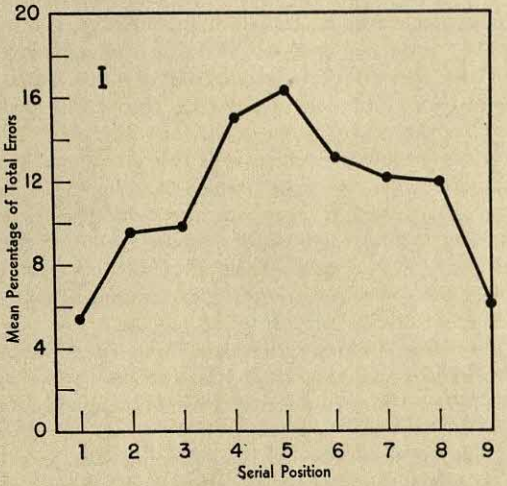


FIG. 1. THE SERIAL-POSITION CURVE IN EXPERIMENT I
The conditions were strictly spatial, *non*-temporal presentation of the serial list, with reconstruction as *S*'s mode of response.

Since we are primarily interested in the *shape* of the serial-position curve, rather than its absolute position on the ordinate, the curves for all these experiments are based on the percentages of total errors that occur at each position. Thus the area under all the curves is the same. The percentage of total errors at each position was determined for each *S* individually, and these percentages were then averaged for the total group. The serial-position curve thus obtained for the data of Experiment I is shown in Fig. 1. The curve is quite typical of serial-position curves obtained by temporal serial presentation of the items. This curve is however, somewhat less skewed, although the greater proportion of errors still occurs beyond the middle position.

EXPERIMENT II

Apparatus. The apparatus used in this and the following experiments is described in the Apparatus Section of this JOURNAL.[3] The aspects of the apparatus essential to this experiment were: (a) the stimulus-display unit, and (b) the response-unit. On a table directly in front of *S* was the response-unit, consisting, in this experiment, of a straight row of nine buttons. The buttons, ⅝ in. in diameter, were evenly spaced 1½ in. apart, measured from the center of the buttons. All the buttons were identical and had no labels or other distinguishing cues. Directly behind the response-unit, in full view of *S*, was the stimulus-display unit, consisting of a 3 × 4 in. screen on which the stimuli appeared. The stimuli were nine brightly colored geometric figures which appeared one at a time in the center of the screen. The figures, approximately 2 in. in size, subjectively appear about equal in size. There were three forms in each of three colors—triangles (T), squares (Sq), and circles (C) colored red (R), yellow (Y), and blue (B). The stimuli appeared on the center of the screen one at a time in an order which throughout the experiment was *random,* with two exceptions: (a) each stimulus-object appeared once within every set of nine stimuli; (b) stimuli of the same shape or color never appeared in immediate succession. The first restriction on the randomness was imposed to insure equal exposure to all stimuli; the second restriction was imposed to make the presentation of stimuli consistent with the rules for presentation in the control experiments (Experiments III and IV), except for the variations that are directly germane to the purpose of the experiments.

Procedure. Twenty-five *S*s served in the experiment. The *S* was instructed that his task was to learn by trial and error which button on the response-panel corresponds to each stimulus. When *S* pressed the correct button, he was reinforced by a 'bong,' following which the next stimulus appeared. If the wrong button was pressed, there was no 'bong,' but the next stimulus would appear. The *S* could not correct a mistake, since the instant one button was pressed, all others went 'dead,' so that pressing the correct button after an incorrect one had been pressed would produce no 'bong.' The rate of presentation of stimuli was governed by *S, i.e.* a new stimulus would not appear until *S* pressed a button, whether the right one or not. The stimulus would remain in view until the button was pressed, whereupon, after a 1-sec. delay, the next stimulus would appear. The purpose of this brief delay was to give *S* time to notice the *S–R* relationship. Also, the reinforcing 'bong' occurred while the stimulus was still in view, but the 'bong' never 'overlapped' the presentation of the next stimulus.

The response-buttons were arranged in a straight row such that the order of the stimuli to which they corresponded was the same as the serial order of the stimuli used in Experiment III, which was intended as the control for comparative purposes. Going from the *S*'s left to right, the stimulus corresponding to each button was: BSq, YT, RSq, BC, RT, YC, BT, YSq, RC.

Throughout the experiment *E* recorded *S*'s responses in *sets of nine.* The *S* had to persist in the task until he attained the criterion of nine correct responses in succession all within a *set.* The criterion was established in this manner in order to have it correspond to the criterion in the other experiments, *i.e.* correct reproduction (or anticipation) of the serial list of nine items.

---

[3] *Infra.,* 470-477.

Results. This trial-and-error learning was by far the most difficult of all the tasks in the four experiments. To attain criterion, *S*s required an average of 298.84 stimulus-presentations; they averaged 161.36 errors, and the mean percentage of errors was 53.98.

There was no systematic pattern in the number of errors or of correct responses made on each of the nine buttons. Individual *S*s seemed to have preferences for particular buttons and used these much more than the others; these preferences may have been determined partly by which buttons happened by chance to give *S* his first few reinforcements. For the total group, however, these positional preferences averaged out, so that all the buttons received an approximately equal number of responses. The slight, unsystematic, and non-significant differences between the total responses for each button do not in any way resemble the usual serial-position curve.

But when the *percentage* of incorrect (*i.e.* non-reinforced) responses is computed for each button, thus ruling out the effect of absolute frequency of use of the button, a serial-position effect becomes evident. In other words, the terminal buttons are learned most readily and those in the middle are learned with the greatest difficulty. To compute the points for this curve properly, the percentage of errors for each button for each *S* must first be determined. Then, so as to weight each *S* equally in the totals, the percentage of errors on each button must be divided by the total of these percentages for all nine buttons; this is done for each *S*. These percentages are then averaged over all *S*s for each button. The result is shown in Fig. 2. One point must be made clear: this curve does not represent the group's positional preferences for different buttons *per se*. It represents the relative difficulty of learning to match the buttons with their corresponding stimuli.

The data of this curve were subjected to an analysis of variance, which showed that the mean percentage of errors is significantly different for the different positions of the buttons. For the *Between Positions* main effect, $F = 3.11$, $df = 8/192$, $p < 0.01$. The most likely reason that the serial-position curve in Experiment II is so flat as compared with the other serial curves is that in the early stages of the trial-and-error learning the errors for the whole group were distributed more or less evenly over all the buttons. Such an admixture of 'random errors' in any set of serial-learning data would produce a general flattening of the serial-position curve. Also, there was greater variability in the serial-position curves of individual *S*s, due to the fact that, regardless of serial position, the button on which *S* chanced to receive his first reinforcement tended to be learned more readily than the other buttons. The fact that in trial-and-error learning *S* always hit upon his first success by chance introduced a good deal of 'random error' into the results, and therefore the serial-position curve could not possibly emerge so conspicuously as it did in all the other experiments. There is no doubt, however, that a serial-position curve bearing the essential features was produced by a *non*-serial presentation of the stimuli and a purely spatial serial arrangement of the response-alternatives.

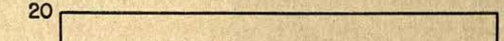


FIG. 2. THE SERIAL-POSITION CURVE IN EXPERIMENT II
The conditions were temporal, *non*-serial presentation of the stimuli, with serial spatial arrangement of the response-alternatives, and trial-and-error learning as *S*'s mode of response. See text for method of deriving the curve.

EXPERIMENT III

Procedure. Sixty *S*s served in the experiment. The stimuli were presented on the screen, automatically paced by the machine at a 3-sec. rate, with a 6-sec. intertrial interval between each set of nine stimuli. The stimuli were presented repeatedly in a constant serial order. The series was preceded by a small white dot in the center of the screen which served as a signal to *S* to anticipate the first item in the serial list.

S was instructed to learn the list by the method of anticipation, responding by saying 'blue square,' 'yellow triangle,' etc., and was encouraged to guess when in doubt. He was also told what stimulus-items would appear—triangles, squares, and circles colored red, blue, and yellow—and that items of the same shape or of the same color would never appear in immediate succession. The order of the stimuli was the same as the order of the response-buttons in Experiment II, viz. BSq, YT, RSq, BC, RT, YC, BT, YSq, RC. *S* began responding on the very first trial, and was required to attain the criterion of one perfect trial, *i.e.* anticipating correctly in immediate succession every item in the serial list.

Results. This experiment, which followed the usual procedure of experiments on serial learning, viz. that of presenting the items in one location in a *temporal* sequence at a constant rate, serves as a basis for comparison with the results of Experiment II, which was a strictly *spatial* type of serial learning.

Since the percentage of errors in Experiment II was 53.98 and in Experiment III was only 44.22, it is evident that *S*s in Experiment III were closer to the asymptote of complete mastery of the task than were *S*s in Experiment II, even though both groups had attained the criterion of nine successive correct responses. The degree of bowing of the serial-position curve is in part a function of the stage of learning; it becomes increasingly peaked as *S* approaches the asymptote of mastery of the list. In view of this fact, it was decided that the serial-position curves produced in Experiment II and Experiment III should be compared, not on the basis of both groups having attained the same predetermined criterion of nine successive correct responses, but rather on the basis of the groups having attained the same degree of mastery of the task as represented by the percentage of errors. In other words, though fewer errors are made in the later stages of learning, they occur increasingly in the middle positions. Thus, while the total percentage of errors continues to decrease throughout learning, the greatest relative increase in errors is in the middle positions. Therefore, it seems reasonable to compare the serial-position curves of the two groups based on the data up to the largest number of trials on which both groups have approximately the same percentage of errors.

Since Experiment III produced the smaller percentage of errors, it was possible to find a point on the learning curve for the group at which they arrived at approximately the same percentage of errors as was reached in Experiment II by the time the criterion was attained. Inspection of the data of Experiment III revealed that the group attained approximately the same percentage of errors as Experiment II at a point that included three-fourths of the trials to the criterion. Therefore, the first three-fourths of the trials for each *S* were used in determining the serial-position curve. On this basis, the percentage of errors in Experiment III was 54.08, as compared with 53.98 in Experiment II. Thus, the serial-position curves for the two experiments are based on almost exactly the same level of mastery of the tasks in terms of the percentage of errors.

The serial-position curves in Experiment III, shown in Fig. 3, were determined in the same manner as in the other experiments. The solid line in Fig. 3 is the serial-position curve based on the first three-fourths of the trials, representing the same error percentage as in Experiment II. The broken line in Fig. 3 is the serial-position curve for all the trials up to the criterion on nine successive correct trials. It is, of course, more peaked. As would be expected, the bowed curves are typical of serial-position curves, with the greater proportion of errors occurring past the middle position in the series.

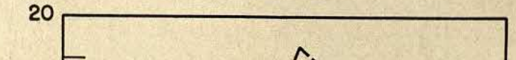


FIG. 3. THE SERIAL-POSITION CURVE IN EXPERIMENT III
The conditions were temporal serial presentation of the stimuli (machine-paced) and verbal anticipation as the mode of S's response. Solid line based on first three-fourths of trials in Experiment II; Broken line, all trials in Experiment III.

EXPERIMENT IV

Procedure. The procedure in this experiment, in which 60 *S*s were used, was exactly the same as in Experiment III, with two exceptions: (a) the rate of presentation of the stimuli was governed by *S;* each stimulus in the series appeared only after *S* had anticipated it. Thus, *S* was under no time-pressure and could take as long as he pleased in making each response. When *S* was in doubt as to the next anticipation, he was forced to guess, since the next item would never appear until *S* made a response. (b) Twenty different orders of stimuli were used, that each order was represented only three times in the over-all results. This procedure would tend to randomize any systematic interaction between the particular stimulus-items (or their arrangement) and their position in the series.

Results. The average number of stimulus-presentations was only 27.14, the average number of errors was 8.83, and the average percentage of errors was 32.52. These figures are remarkably close to those of Experiment I, which were 26.70, 9.06, and 33.96, respectively. The *temporal* serial task appears to be of the same difficulty as the *spatial* serial task when *S* is allowed to respond at his own speed.

Since *S* began responding in Experiment IV on the very first trial without previous exposure of the serial order, the task was not comparable to Experiment I in this respect. In Experiment I *S* viewed the serial order for 10 sec. before attempting to reconstruct it on the first trial. Also, since so few trails were required to attain the criterion in Experiment IV, the first trial, which represents sheer guessing, constitutes a relatively large proportion of the total errors. Of course, the effect of serial position cannot emerge on the first trial when it is based on sheer guessing. Therefore, the serial-position curve, if it is to be properly compared with that of Experiment I, should be based only on all the errors made *after* the first trial. This was done, and the serial-position curves, shown in Fig. 4, were determined in the same manner as in the other experiments. The solid line is the curve of percentages of errors excluding the first trial. The broken line shows

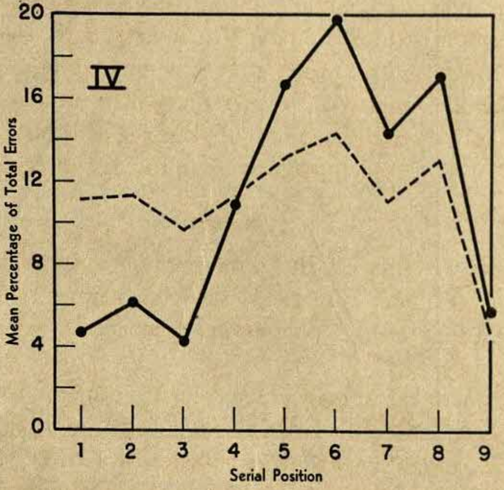


FIG. 4. THE SERIAL-POSITION CURVE IN EXPERIMENT IV
The conditions were temporal serial presentation of the stimuli (*S*-paced) and verbal anticipation as the mode of *S*'s response. Solid line, percentage of errors excluding the first trial; Broken line, percentage of errors for all trials.

the percentage of errors for *all* trials. Of course, this curve is less peaked, since the errors on the first trial were more or less evenly distributed over all the positions, with the exception of the last position, which seems to show a marked 'recency' effect. Actually, of course, this low percentage of errors in the last position, especially for the first trial, is not a true 'recency' effect, but must be due to the fact that when *S* is not under time-pressure he can fairly well remember which item has not been used in the list up to the last item, which makes it easy for *S* to 'guess' correctly the last item on his very first time through the series. Another interesting feature of the solid curve is the uniformly low percentage of errors in the first three positions, which indicates that most *S*s were able to master the sequence of the first three items after just one exposure. Without time-pressure, especially, one would expect immediate memory-span to play a part in learning the serial list, and it appears that this factor is represented by the equally low percentages of errors in the first three positions.

SUMMARY

The essential features of the serial-position curve have been shown clearly to emerge under conditions involving neither serial anticipation, nor temporal order of the stimulus-items, nor even serial order of the stimulus-items.

In Experiment I, the stimuli (nine colored geometric forms) were presented for 10 sec. on each trial in a *spatial* serial arrangement (a straight row), all the stimuli appearing simultaneously. After each 10-sec. exposure, *S* had to reconstruct, without time-limit, the serial arrangement. The errors made in reconstruction were distributed in the typical bowed curve characteristic of serial-position curves produced under the usual conditions (represented in Experiments III and IV) of temporal serial presentation with verbal anticipation as the mode of response.

In Experiment II, the stimuli (nine colored geometric forms) were presented temporally, at a *S*-paced rate, in a *random* order. The *S* had to learn to match each of a set of nine identical buttons, arranged in a row, with each of the nine stimuli. Reinforcement for pressing the button corresponding to the stimulus appearing on the screen consisted of a 'bong.' Thus, *S*'s mode of response was trial-and-error selective learning. The only *serial* aspect of the experiment was the *spatial* serial arrangement of the response-alternatives. When the percentage of errors made on each button for all trials up to the criterion of nine successive correct responses (the same criterion as in the other experiments) was determined, it was found that the relative difficulty of learning the buttons varied as a function of their position, the easiest to learn being those at the ends and the most difficult being those in the middle. In other words, the pattern of errors had the essential features of the usual serial-position curve. It did not, however, manifest the typical skewness of most serial-position curves. The distribution of the percentages of errors on the various buttons does not reflect positional preferences *per se*. While individual *S*s showed positional preferences, reflected in the different numbers of total responses on each of the buttons, these preferences did not correspond to the serial-position curve, nor were they characteristic of the group as a whole. The total responses on each button for the group were nearly equal.

Experiments III and IV were intended as control experiments; they involved the same type of stimuli and general procedures as the other experiments, except that they maintained the essential features of temporal serial presentation and verbal anticipation on which *S–R* theories of serial rote-learning are based. In Experiment III, the stimuli were machine-paced as a 3-sec. rate, with a 6-sec. intertrial interval; in Experiment IV, the stimuli were *S*-paced, each stimulus appearing only after *S* had responded.

These findings would seem to have significance particularly for those *S–R* theories of the serial-position effect which depend upon a temporal sequence of the items and serial anticipation by *S*. The conditions of Experiments I and II, for example, completely lack the features which would permit one to invoke Lepley's hypothesis of 'trace-conditioned responses' or Hull's hypothesis of 'inhibition of delay' and its corollary that greater amounts of inhibitory potential span the middle items of the list.[4] The same can be said of Hull's theory of serial learning in 1952, which is based on the 'gradient of reinforcement,' *i.e.* the effectiveness of reinforcement in establishing an associative bond is assumed to decrease as a function of the delay in the reinforcement.[5] Yet, in Experiments I and II the conditions necessary for the operation of the variables postulated by Hull to explain the serial-position effect clearly do not exist. Still, however, the typical bowed serial-position curve comes forth. Apparently these theories of the serial-position effect either are incorrect or are not sufficiently general to explain the emergence of the bowed serial-position curve under conditions other than the learning of a temporal sequence by the method of serial anticipation.

---

[4] W. M. Lepley, Serial reactions considered as conditioned reactions, *Psychol. Monogr.*, 46, 1934, No. 205, 40-45; C. L. Hull, *et al.*, *Mathematico-Deductive Theory of Rote Learning*, 1940, 110, 175-177.

[5] Hull, *A Behavior System*, 1952, 156-191.

TEMPORAL DIFFERENTIAL SENSITIVITY FOR AUDITORY STIMULI

By ARNOLD M. SMALL, JR., and RICHARD A. CAMPBELL,
University of Iowa

As Woodrow has pointed out, temporal stimuli are generally of two sorts: (1) empty intervals—bounded by brief stimuli, and (2) filled intervals—defined by a continuously present stimulus.[1] The response-tasks, moreover, have traditionally been divided in terms of whether intervals are (a) compared or (b) reproduced. The present experiment is one in which filled intervals were compared.

Specifically, the purpose of the present study was to determine the differential threshold for duration, using a wide range of acoustic stimuli. It was especially desired to investigate discrimination at very short durations.

METHOD

Apparatus. Electronic switches (Grason-Stadler, 829574) and associated timers (Grason-Stadler, E3299A) were arranged in such a way that four independently variable time-intervals were generated. The first interval contained the standard stimulus (*STD*); the second was an interval of silence—the inter-stimulus interval (*ISI*); the third interval contained the variable stimulus (*VAR*); and the fourth interval was one of silence separating successive pairs of standard and variable stimuli. This fourth interval was maintained at 5 sec. throughout the experiment. Interval-durations were monitored with an electronic digital counter (Berkeley, 5500).

An audio oscillator (General Radio, 1304B) and noise generator (Grason-Stadler, 455B) together with appropriate attenuators furnished the standard and variable stimuli for different parts of the experiment. The circuits were so arranged that the stimuli would be gated on at a zero-crossing of the electrical stimulus-wave form, thus the voltage discontinuities and associated transients in the acoustic wave form were minimized. Whether the stimulus is gated off at a zero crossing is a function of duration and frequency of the stimuli; consequently, values of these parameters for the standard were so chosen as to maximize this event. The variable was also gated on at a zero-crossing, but no attempt was made to gate it off in a similar manner.

These stimuli, together with a calibration signal, were recorded on magnetic tape (Scotch 111-24R). During the experimental session, the tapes were played back at 7.5-in. per sec. with the same recorder (Ampex 350). The output of the re-

---

* Received for publication January 21, 1961. This research was supported in part by Grant G-10606 of the National Science Foundation.

[1] Herbert Woodrow, Time perception, in S. S. Stevens (ed.), *Handbook of Experimental Psychology,* 1951, 1224.

corder was led through a power amplifier, attenuator, matching transformer and into 15 pairs of dynamic earphones (Permoflux PDR-8 in MX 41/AR cushions). Every standard stimulus was presented at a constant loudness-level of 50 phons. This was determined by a preliminary alternate binaural loudness-match against a 1000∼ tone. Table I gives the actual sound-pressure level (*SPL*) generated in a 6-cc. coupler for the stimuli presented.

Procedure. The method of constant stimulus-differences was used, with the *O*s' being instructed to judge whether the second stimulus of the pair was *longer* or *shorter* than the first.[2] For a given duration of the standard stimulus, the variable could, with equal liklihood, be any of seven durations—ranging from 0.25 to 2.5 times the standard. Each *O* contributed 84 longer-shorter judgments (distributed equally among the 7 variable durations) for each condition.

The first of the 13 one-hour experimental sessions was devoted to screening, instructions, orientation, and practice. In each of the next 12 sessions, the *O*s were presented a standard (*STD*) consisting of a single duration-frequency combination but with all inter-stimulus intervals (*ISI*). Stimulus-pairs involving the same *ISI* were grouped together, but the order of *ISI*-presentation was random for each session. The *O*s were divided into four groups, nearly equal in number, each of which received different random orders of *STD* duration-frequency combinations.

TABLE I

SOUND-PRESSURE-LEVEL (db., re: 0.0002 μbar) OF STIMULI USED
All table entries were judged equally loud in a preliminary experiment.

Stimulus	Duration of standard stimulus (m.sec.)			
	0.4	4	40	400
250∼	93	93	86	81
5000∼	101	95	87	81
noise	106	94	86	82

The experiment was carried out in a sound-treated room where groups of about 12 *O*s were seated three to a table and separated by large plywood partitions. The *O*s marked their responses on a suitably coded score-sheet and subsequently each sheet was machine-scored and these scores entered on punched cards for further analysis.

Observers. The *O*s (47 in number), were students from a course in introductory psychology. The only criterion for their selection was that they had no hearing loss in either ear as determined by standard audiometric procedure.

RESULTS

Differential limens. The differential limens (*DL*s) were determined graphically. The probability of a longer judgment for each stimulus-condition was plotted on probability paper and a straight line fitted visually.[3] It was noted that at extreme values of the variable stimuli, the data sometimes deviated markedly from a cumulative Gaussian distribution. When this occurred, only the middle five values of the variable stimuli were used in fitting the straight line. The *DL* was defined as one-half the difference

---

[2] J. P. Guilford, *Psychometric Methods*, 1936, 186-200.

[3] Guilford, *op. cit.*, 183.

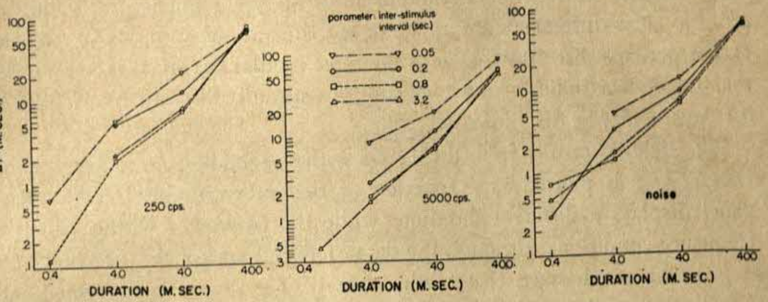


FIG. 1. *DL*s FOR DURATION (ΔT) AS A FUNCTION OF DURATION, INTER-STIMULUS-INTERVAL, AND FREQUENCY

Every point represents 3948 judgments.

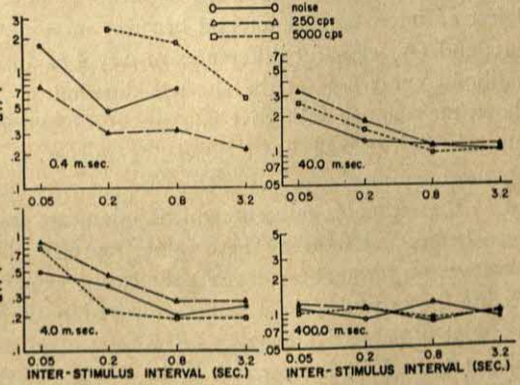


FIG. 2. RELATIVE *DL*s FOR DURATION ($\Delta T/T$) AS A FUNCTION OF DURATION, INTER-STIMULUS INTERVAL, AND FREQUENCY

Every point represents 1008 judgments. The vertical scale ($\Delta T/T$) differs slightly among the groups.

between the stimulus-value yielding 25% longer judgments and the one yielding 75% longer judgments.

Figs. 1 and 2 show the *DL*s as determined in the present study. In Fig. 1, the just-discriminable change in duration (ΔT) is plotted as a function of the standard duration. Each graph shows a different frequency condition while the parameter within them is the inter-stimulus interval.[4] If Weber's Law held, all curves in Fig. 1 should have a slope of unity. Instead, the slope is generally less than unity and tends to become markedly less as duration is shortened. It can also be seen that there is an ordering such that the shortest inter-stimulus intervals yield the largest values of ΔT. No clear differences are apparent for the various stimulus-frequencies except perhaps that 5000~ yields the most regular curves. It is to be noted that these data indicate that two auditory stimuli differing by as little as 0.12 m.sec. may be discriminated.

Fig. 2 shows results from the 12 *O*s with the smallest *DL*s. The relative $DL(\Delta T/T)$ is plotted as a function of the inter-stimulus interval. Each panel displays a different duration, while the parameter within panels is frequency-condition. Although the data of Fig. 2 are for the 12 "best" *O*s, it is to be emphasized that if the data of Fig. 1 are transformed into ΔT/T, the resulting plot would be similar to Fig. 2. The curves of the 'best' *O*s are simply displaced downward from the position which would be occupied by curves derived from the *O*s. Three things should especially be noted in Fig. 2: (1) $\Delta T/T$ increased markedly as duration shortens; (2) The effect of inter-stimulus interval becomes more obvious as duration shortens; and (3) clear-cut differences in $\Delta T/T$ as a function of frequency condition appear only at the shortest duration.

Fig. 3 shows the point of subjective equality as a function of the inter-stimulus interval. Each panel is for a different standard stimulus-duration and the parameter within the panels is the frequency-condition. The *PSE* was defined as that stimulus-value at which judgments were evenly divided between longer and shorter. These values represent the mean of all 47*O*s. If there was no constant error (*CE*) the curves would be horizontal lines intersecting the vertical scale at values indicated by the panel legends (that is, the point of subjective equality (*PSE*) would equal the point of object equality (*POE*)). It is clear that if the *PSE* is less than the *POE*, then the *O*s are underestimating the duration of the standard stimulus; the converse, of course, is also true. It would appear from Fig. 3 that for a standard stimulus of 400 m.sec. there is no *CE*. At 40 m.sec. there is an indication of a positive *CE* (overestimation of the standard) at only the shortest inter-stimulus interval. For a 4-m.sec. standard stimulus, there is a trend for a negative *CE* at short inter-stimulus intervals (*ISI*) with a possibility of a slight positive *CE* at the longest values of *ISI*. The existing data at 0.4 m.sec. suggests a similarity to that of the 4-m.sec. duration.

---

[4] Both stimulus-duration and inter-stimulus interval were specified in terms of the voltage across the earphone. This does not correspond exactly to the acoustical output of the earphone, but has the advantage of relative unambiguity in specification of starting and stopping time. The characteristics of the circuitry and earphones used were such that the acoustic signal continued in a exponentially decaying fashion some 0.5 to 0.75 m.sec. (measured at the point where the amplitude was 10% of the original amplitude) after termination of the electrical signal.

It should be pointed out that a few points of the data are not plotted in Figs. 1-3. In those cases the psychometric functions had either a slope of essentially zero, or a scatter of points such that a fitted straight line had no real descriptive meaning.

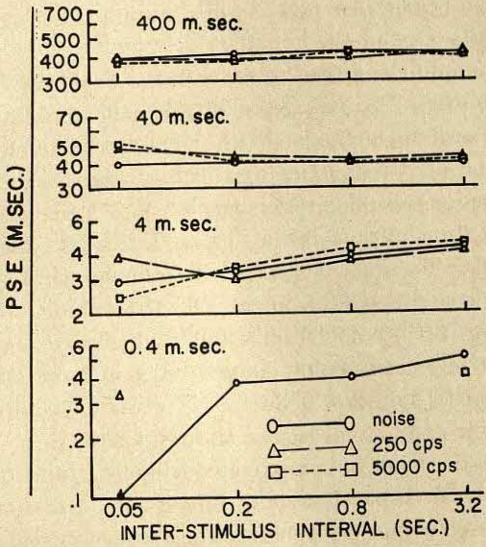


FIG. 3. *PSE*s FOR DURATION AS A FUNCTION OF DURATION, INTER-STIMULUS INTERVAL, AND FREQUENCY
Each point represents 3948 judgments.

Analysis of variance. Three-way analyses of variance (duration, inter-stimulus interval, and frequency) were performed on the data derived from *DL*s (Fig. 1) and *PSE*s (Fig. 3). Because of the manner in which the psychometric functions were formed, no within treatments mean square was available. Consequently, it was assumed that the triple interaction was not significant, and it was used as the error-term.[5] Relative measures were used, $\Delta T/T$ and *PSE/POE*, to offset the obvious effects of duration. For instance, the *PSE* nearly always approximates the *POE;* therefore, if the duration of the standard stimulus is changed grossly, it would be expected that the *PSE* would also change. With this particular method of analysis the only significant source of variance is the *duration* of the standard stimulus.

---

[5] A. L. Edwards, *Experimental Design in Psychological Research*, 1950, 252-260.

DISCUSSION

The effect of acoustic stimuli. At very short durations there are certain characteristics of the stimuli that may contribute importantly to the discrimination of auditory durations. One such characteristic is the 'ringing'-time of the earphones. One of the important determiners of ringing-time is the instantaneous amplitude of the acoustic waveform at the moment the electrical signal terminates. To a first approximation, ringing-time is proportional to 'final instantaneous amplitude.' With pure tones as stimuli for *STD* durations of 40 and 400 m.sec., ringing-effects are so small (less than a 5% increase in duration) that they may be neglected.

At standard stimulus-durations of 4 m.sec., however, the effect becomes important. For all durations of the variable stimulus (*VAR*) with 5000~, the *VAR* terminated at a zero crossing, hence, minimal ringing. For the 250~-stimulus, 'final instantaneous amplitude' varied as a function of *VAR* duration.[6] This uncertainty (variability) or *VAR* duration leads to the prediction that the *DL*s obtained for 250~ should be larger than those for 5000~. Fig. 2 shows this to be the case.

At 0.4 m.sec. for the pure tones an additional factor must be considered. Although the peak amplitudes of the sine waves constituting the standard stimuli were all adjusted for equal loudness, the peak amplitude of each variable was always the same as its corresponding standard. Thus, the 250~-tone, although again contributing durational uncertainty through ringing-time, provided another cue for discrimination. This cue was that of an intensive variation as a function of duration. Variable durations ranged from 0.1 to 1.0 m.sec.; corresponding to phase-angles of 9° to 90° of a 250~ sine wave. The amplitude of a sine wave varies about 16 db. over that phase angle range. Clearly, with an additional cue of this magnitude available it should not be surprising to note that the 250~-stimulus shows smaller *DL*s at 0.4 m.sec. duration than does 5000~ (Fig. 2).

It might appear as though noise were an ideal standard against which to compare these short durational characteristics of pure tones. Unfortunately, however, noise exhibits certain unique effects because of its random nature. For wide-band noise the distribution of instantaneous amplitudes is approximately Gaussian with a mean of zero.[7] Some variation in the duration of noise-stimuli will, therefore, be introduced by variation in ringing-time.

---

[6] The effect of lower frequency deals with the fact that in 4 m.sec. a 5000~ tone goes through 20 vibrations while a 250~ tone goes through only 1. Thus, for 250~ if the 4 m.sec. period is sectioned at three points in time, the instantaneous amplitudes at several of these points differ.

In addition, each noise-stimulus, either *STD* or *VAR*, constitutes a sample of several pseudo-random variables (power and spectral character). Since the samples are drawn from the same 'population' (noise-source), samples will become more similar as the sample-size (duration) is increased. Conversely, the shorter the sample, the larger the sampling variations. If power and spectral character are important in discriminating duration, it might be expected that noise at short durations would exhibit larger *DL*s than would be found with a 'constant stimulus.' The results shown in Fig. 2 for 4.0 m.sec., tend to support this argument, but the reversal between noise and 5000~ for 0.4 m.sec. would not be predicted.

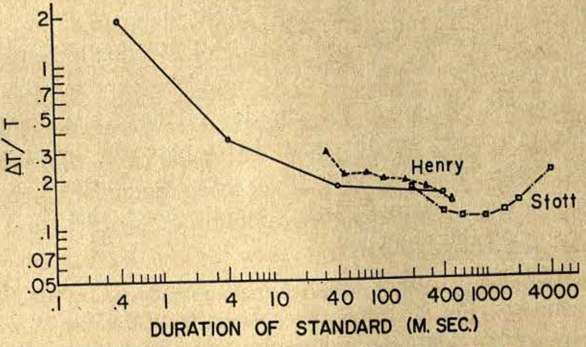


FIG. 4. COMPARISON OF PRESENT RESULTS WITH THOSE OF OTHER INVESTIGATORS

At the very short durations then it would seem that *DL*s and to a lesser extent the *PSE*s are influenced by characteristics of the stimuli and consequently do not reflect in a simple fashion the *O*'s ability to discriminate duration.

Comparison to other studies. Because of differences in method and stimuli in the present study, no directly comparable data exist. Fig. 4, however, compares a sample of the present data to the results of two previous studies. Henry's data are for a tone of 500~, 50 db. above threshold, and unspecified inter-stimulus intervals.[8] The data of Stott are for a 1000~

---

[7] J. H. Laning, Jr. and R. H. Battin, *Random Processes in Automatic Control,* 1956, 155-60.

[8] F. M. Henry, Discrimination of the duration of a sound, *J. exp. Psychol.,* 38, 1948, 734-743.

tone at an unspecified level with an inter-stimulus interval of 1.5 sec.[9]. For those durations in which the data from various studies overlap, there is fair agreement. This would probably remain true regardless of which condition had been chosen to represent the present study, since for long durations the other variables have little effect upon the relative *DL* (see Fig. 2).

Loudness as a basis of temporal discrimination. There has been a consistent attempt in the past to explain changes in $\Delta T/T$ as a function of duration in terms of concomitant loudness-changes.[10] That is, it has been reported that as duration is shortened, the loudness of the stimulus decreases[11] and there is evidence that as intensity (loudness) is decreased the *DL* for intensity (loudness) increases.[12] If the standard stimuli were presented with constant intensity, and if duration is actually being discriminated on the basis of loudness, then it would be expected that the *DL*s for duration would increase as the duration of the standard stimulus decreases. This, of course, was the general trend seen in the present study. In this experiment, however, sound-levels were so adjusted that all standard stimuli had equal loudness. It is clear, therefore, that the current results do not support this argument.

Although the standard stimuli were equated in loudness, the loudness of the variable stimulus was free to change with changes in its duration. The decrement in loudness for a given decrement in duration is given by the slope of the function relating loudness to duration. Since most studies indicate that the slope of the function relating loudness (phon) to long duration becomes steeper as duration is shortened, a given relative change in duration (for instance, a 5:1 change) would result in a greater increment of loudness for a stimulus of short duration than for one of longer duration.[13] If loudness were a primary cue, then, it might be expected that the relative *DL* for duration would become smaller as the duration is shortened. To the contrary, however, the data of the present study indicate the relative *DL*s increase, generally, as the duration is decreased. It should be noted that at least two studies report loudness (phon)-log duration functions which, over the range of durations used here, show a constant slope.[14] Indeed, when loudness-level in phons is converted to loudness in sones, the function's slope actually decreases as log-duration decreases—a condition sufficient to predict the present data. It would seem, therefore, that a clarification of the loudness-duration function must necessarily precede the use of that function as an explanation of temporal discrimination.

---

[9] L. H. Stott, Time-order errors in the discrimination of short tonal durations, *J. exp. Psychol.*, 18, 1935, 741-766.

[10] Henry, *op. cit.*, 740-742.

[11] See, for instance, George von Békésy, *Experiments in Hearing*, 1960, 236-238; W. R. Garner, The loudness and loudness matching of short tones, *J. acoust. Soc. Amer.*, 21, 1949, 398-403; W. A. Munson, The growth of auditory sensation, *ibid.*, 19, 1947, 584-591.

[12] R. R. Riesz, Differential intensity sensitivity of the ear for pure tones, *Phys. Rev.*, 31, 1928, 867-875.

[13] Garner, *op. cit.*, 403; Von Békésy, *op. cit.*, 238. Munson, *op. cit.*, 585.

[14] G. A. Miller, The perception of short bursts of noise, *J. acoust. Soc. Amer.*, 20, 1948, 160-170; A. M. Small, Jr., J. F. Brandt, and P. G. Cox, Loudness as a function of signal duration, *J. acoust. Soc. Amer.*, 34, 1962, 513-514.

An alternate hypothesis might postulate that the standard stimulus reduced the loudness of the variable which followed it. Thus, if the relative *DL* for 'loudness' increases as loudness decreases and if it can be further assumed that this decrement in loudness increases as the duration and inter-stimulus interval are shortened, predictions based on this hypothesis would be in line with the data of the present study. If, however, the standard influenced the loudness of the variable and loudness was the only cue for temporal discrimination, then this should be reflected as a substantial *CE* in the *PSE*. This hypothesis appears untenable since the *PSE* as determined in the present study does not show the predicted changes (see Fig. 3).[15]

Perhaps the simplest hypothesis which retains loudness as the primary cue is that as duration of the stimuli shortens, maintaining loudness constant, there is a deterioration in the discrimination of loudness. There seem to be no data available on this point; clearly further research is necessary.

Clock-mechanisms as a basis of temporal discrimination. In terms of simplicity, a clock-mechanism of time-perception has historically had great appeal.[16] In the context of the present study, the mechanism usually proposed consists of three parts: (1) a constant time-base—perhaps in the form of equa-spaced pulses; (2) a device which allows these pulses to pass to the third device only when the stimulus is audible; and (3) a mechanism which counts the pulses. It is difficult to reconcile the present data with predictions based on this mechanism. For instance, neither of the two simplest assumptions regarding criteria for discrimination, an absolute or relative difference in count between two stimuli, seem tenable.

This counting model has, recently, been modified and incorporated into an hypothesis which views *O* as a tester of statistical hypotheses.[17] The pulses which form the time-base are assumed to be derived from a large number of independent elements whose time of firing is randomly distributed. Further, because of background noise, either internally or externally generated, there would be uncertainty as to the exact starting and stopping time of the stimulus and hence uncertainty in number of counts. Finally, an imperfect memory is assumed in comparing the number of counts in one interval to those in another. This model makes no specific prediction with respect to the frequency of the stimulus, but for a given inter-stimulus interval it predicts quite well the size of the *DL* as a function of duration. Unfortunately, the prediction as to change in *DL* as a function of inter-stimulus interval is in a direction opposite to that shown in the present data.

---

[15] The predicted change would be such that the variable stimulus would be set greater than the standard for subjective equality.

[16] For a recent review of this literature on temporal discrimination see C. D. Creelman, Human discrimination of auditory duration, Technical report No. 114, Cooley Electronics Laboratory, University of Michigan, 1960.

[17] Creelman, *op. cit.*, 38-52.

Time-order error. Time-order error (*TOE*) has, of course, been the subject of extensive investigation. Several studies have dealt with *TOE* resulting from the discrimination of auditorially defined temporal intervals and nearly all found it in substantial amount.[18] Yet in the present study *TOE* is distinguished principally by its relative absence. *TOE*, when it was observed, occurred at the inter-stimulus interval of 0.2 sec. or less and changed sign as a function of duration. This is an interesting effect whose explanation is not *immediately* apparent. Although this experiment superficially resembles others previously reported, considerable difference exists in procedure, stimuli, and *O*s. The present results should not, therefore, be construed as necessarily contradictory.

SUMMARY

This study sought to gain information regarding the discrimination of auditorially presented temporal intervals for various short durations. In addition, inter-stimulus intervals and frequency were varied. Using the method of *constant stimulus-differences,* we found that temporal discrimination deteriorated as duration was shortened. As the inter-stimulus interval was shortened, discrimination also became poorer; but inter-stimulus interval lost its influence as duration lengthened. Stimulus-frequency seemed to be an important factor only at the shortest duration. The point of subjective equality generally remained close to the point of objective equality, the only exceptions being at short durations and inter-stimulus intervals.

The results at very short durations are generally compatible with hypotheses based on the nature of the acoustic stimulus. For those conditions which were comparable to those of previous investigators, results were similar. Previously proposed hypotheses based on the premise that loudness or a clock-mechanism is the primary determinant of temporal discrimination do not, however, in their present form, adequately explain the data.

---

[18] See, for instances, S. Natajima, The time-error in the successive comparison of tonal durations, *Jap. J. Psychol.,* 29, 1958, 27-39; Stott, *op. cit.,* 749-750.

STIMULUS-CONDITIONS AFFECTING A RECENTLY DISCOVERED STEREOSCOPIC EFFECT

By BENJAMIN W. WHITE, Massachusetts Institute of Technology

In a recent article, Julesz has described an unusual stereoscopic display consisting of a pair of rectangular dot-matrices such as those shown in Fig. 1.[1] The dots in the matrices differ in brightness, the permissible values being distributed randomly over the matrix. The two matrices are identical in the peripheral region. In the inner region, the columns of the right matrix have been transposed to the left, the left-most column having been brought around to the right edge of the inner matrix. Table I gives a schematic illustration of this operation. Numbers are used to represent the brightness-values of the dots in each position, and the line indicates the boundary between the inner and the outer regions. Note that in the right matrix the left-most column of numbers (3, 4, 8, 1) is identical with the second column of the inner region of the left matrix. For simplicity, this illustration shows small matrices. The matrices shown in the first figure have a peripheral region which is 64 by 64, an inner region which is 32 by 32, two brightness-values, and a shift of one position. If these are presented, one to each eye in a common stereopticon, most *O*s report seeing the inner region as a plane distinctly in front of and separate from the plane of the peripheral outer matrix. If the right and left matrices are exchanged, the impression of depth still is reported, but reversed, the inner matrix being seen behind the peripheral with a rectangular hole in it. The nature of the boundary between the inner and outer regions is especially interesting. The perceived edge of the inner region lies *between* the rows and columns of dots, and there is a sharp edge traversing the black region (indicated by the lines in Table I). This is a rare instance of visual contour-formation in the absence of a brightness-gradient in the field. Most *O*s report the depth readily, an impression which is stable over prolonged exposure. Those *O*s who do not report the effect appear to lack stereoscopic vision because of severe visual impairment in one eye.

This display offers unique advantages in the study of the stimulus-control of

---

* Received for publication February 2, 1961. From the Lincoln Laboratory, operated with support from the U. S. Army, Navy and Air Force.

[1] Bela Julesz, Binocular depth perception of computer-generated patterns, *The Bell System Tech. J.*, 39, 1960, 1125-1162.

stereoscopic depth-perception. First of all, the effect is truly stereoscopic. Closing either eye while viewing the matrices immediately erases all impression of depth, and the contour between the inner and the other regions vanishes. Since the display is abstract in content, the depth-effect cannot be attributed to previous experience with a certain class of objects. Furthermore, most of the other so-called cues for depth are inoperative in this situation: convergence, accommodation, interposition, and linear and aerial perspective. The display has the further advantage that it is easily generated on a computer equipped with a cathode ray tube *(CRT)* and camera, thus making it possible to change stimulus-parameters rapidly and easily.

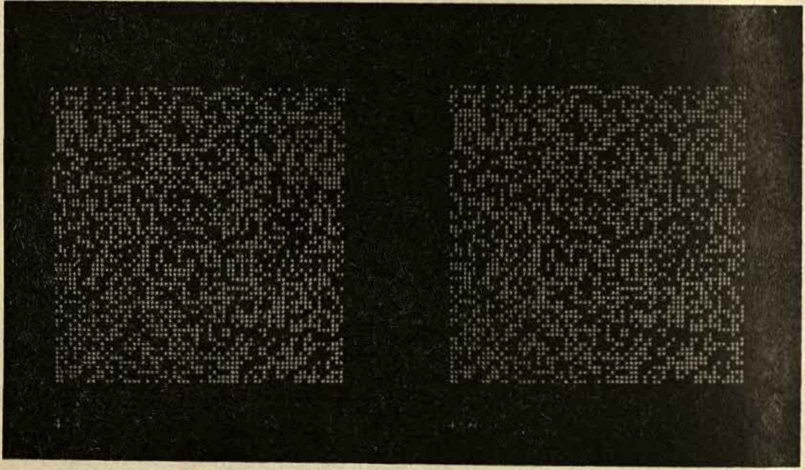


FIG. 1. STEREOSCOPIC DISPLAYS SIMILAR TO THOSE EMPLOYED BY JULESZ

TABLE I

SCHEMATIC REPRESENTATION OF THE STEREOSCOPIC DISPLAY SHOWN IN FIG. 1

Numbers are used to represent the brightness of elements; lines indicate the boundary of the inner region within which transposition takes place.

8	8	6	5	7	2	1	0
4	9	7	6	6	7	4	2
9	1	2	3	1	5	9	5
8	7	1	4	8	7	9	1
2	4	3	8	9	7	3	9
1	5	8	1	4	4	9	6
6	8	0	2	7	5	2	6
4	1	0	5	0	5	8	8

8	8	6	5	7	2	1	0
4	9	7	6	6	7	4	2
9	1	3	1	5	2	9	5
8	7	4	8	7	1	9	1
2	4	8	9	7	3	3	9
1	5	1	4	4	8	9	6
6	8	0	2	7	5	2	6
4	1	0	5	0	5	8	8

The displays employed in this study were generated on the IBM 709 and photographed either with the standard scope-camera or with a Polaroid camera. Most of the dot-matrices were approximately 1 in. square as printed. They were viewed in a Keystone No. 50 Home Training Stereoscope. Since the positioning of the two matrices was not critical, they simply were mounted with rubber bands on a rigid board which could be inserted in the stereoscope. The *O*s adjusted the two prints until binocular fusion of the images was achieved. As Julesz reports, most *O*s, upon first examining such a binocular display, ask what they are supposed to see. Then, after a few seconds' exposure, they exclaim with some surprise: "Now I see it!" The perception of the two planes separated in depth can occur when the *O* has in no way been prepared to see such a thing. Using the perceived width of the inner plane as a unit of distance, most *O*s are able to make some estimate of the separation of the two planes in depth.

Consider two such matrices whose dots can take one of two brightness-values—0 or 1; *i.e.* in every position in the matrices there is no dot at all, or there is a dot of sufficient brightness to be discriminated with ease. With the matrices superimposed, let us subtract the brightness of each position in one matrix from the brightness of the corresponding dot in the other matrix. For the peripheral region, all these differences will be zero since the dots correspond exactly. For the inner region, if the probability of a 1 in any position of either matrix is 0.5, half the differences will be ±1, and half will be 0, on the average, and these differences will be randomly distributed over the inner region. If we now displace one of the matrices to the extent of one column in the proper direction on the *x*-axis, the distribution of differences will change markedly in the inner region. For all but the left-most column of one matrix and the right-most column of the other matrix, the inner region will have zero-differences. In other words, with the rules by which these matrices were constructed, there exists a horizontal shift which will drastically increase the number of zero-differences between corresponding positions in the inner regions of the two matrices. Some such processing of the inputs to the two eyes must take place in the visual system for the stereoscopic effect to occur, and Julesz has described a model based upon such difference-fields.

With such a model of the process underlying this stereoscopic effect, it is possible to think of many ways of altering the display to make such processing difficult, and then to see if these alterations do indeed interfere with the perception of stereoscopic depth. The remainder of this report consists of an account of the various ways in which such displays have been modified, and the degree to which such modifications influence the perception of stereoscopic depth.

(1) Distribution of dot-intensities. If both matrices were composed of dots of identical intensity, the stereoscopic effect obviously would not occur. In terms of the difference-field model, the differences would be zero throughout the matrix with no lateral displacement. What kind of distribution of intensities is sufficient to produce the stereoscopic effect? To study this variable, displays were created containing various numbers of different dot-intensities. It should be noted here that, with the computer-generated displays, intensity is controlled by the number of times a dot is painted. The brightness of a dot certainly is not a linear function of the number of the paints, given a certain film- and lens-setting, and after a certain number of paints, there is no further perceptible increase in intensity. Within these limitations, however, these displays provide some information about the boundary-conditions for this perceptual effect.

As was stated above, the *number* of different dot-intensities in the matrices must be greater than one for the effect to occur. Several displays were generated with two dot-intensities, and it is clear that two intensities are sufficient. In some of these displays, one intensity-value was zero, which meant in effect that either a dot occupied a position in the matrix or the position was unoccupied. The three-dimensional effect was readily elicited under this condition. Another set of displays with 16 brightness-levels randomly distributed with equal probability over the matrices also gave a clear stereoscopic effect.

Given that two different intensities are sufficient to produce the effect, one may ask how different the two intensities need to be. Four pairs of displays were generated, differing only in the separation between the two intensity-levels employed.

TABLE II

COMPUTATION OF 'MIDDLE' BRIGHTNESS-VALUE

	Column No.									
	1	2	3	4	5	6	7	8	9	10
A Sample row of left inner matrix	8	2	2	8	2	8	2	2	8	8
B Same row of right matrix transposed as shown in Table I	2	2	8	2	8	2	2	8	8	8
C Same row of right matrix with each value changed by one unit in direction of neighbor to the right	7	2	3	7	3	7	2	3	8	8
D Same as *C*, but with change of two units in direction of neighbor to the right	6	2	4	6	4	6	2	4	8	8

If intensity is expressed simply by the number of times a particular dot was painted on the scope, the four sets had the following brightness-levels: 4 and 0, 2 and 0, 1 and 0, 2 and 3. In each case, the effect was clearly evident. The last set, with brightness-levels 2 and 3, appears at first glance to be composed of dots of uniform brightness. With this crude technique of brightness-control, it is not too meaningful to extend such a series to the limit. At some point the brightness-difference will be too small to make the stereoscopic effect possible. Knowing whether it is between 15 and 16 paints of the dot, or between 5 and 6, will not aid us in understanding the phenomenon unless there is a satisfactory way of measuring dot-intensity directly. At the moment, it is possible to say only that the intensity-difference does not need to be so large that it is immediately discriminable.

(2) Computed 'middle' brightness-value. A slightly different approach to the question of the brightness-difference sufficient to bring about the stereoscopic effect is illustrated in Table II, where Row *A* represents a particular row in the linear region of the left matrix made up of dots whose intensity-values are 2 or 8 at random with equal probability. Row *B* shows the corresponding row of the inner region of the right matrix under the transposition shown in Table I. In this row, the brightness-values in Columns 1-9 are the same as those in Columns 2-10 in the Row *A*. If we represent the brightness-value of a cell in the inner region of left matrix as $L_{i,j}$ and the corresponding cell in the inner region of the right matrix as $R_{i,j}$, this transposition can be expressed by the equation:

$$R_{i,j} = L_{i+1,j}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots [1]$$

Instead of transposing the values completely, we might merely change them in the direction of the neighboring column by the following rule:

$$\begin{aligned}&\text{if } L_{i,j} = L_{i+1,j},\ R_{i,j} = L_{i,j}\\&\text{if } L_{i,j} > L_{i+1,j},\ R_{i,j} = L_{i,j} - K\\&\text{if } L_{i,j} < L_{i+1,j},\ R_{i,j} = L_{i,j} + K\\&\text{where } 0 < K < L_{i,j} - L_{i+1,j}. \qquad [2]\end{aligned}$$

Such a transformation with a K-value of 1 is shown in Row *C*. Displays generated by this transformation did not yield the stereoscopic effect, while those generated with a K-value of 2, represented in Row *D*, did yield the effect. In terms of the difference-field model suggested above, the fact that the effect can occur under this condition indicates that obtained differences do not have to be zero with a given displacement for the effect to occur. This observation is corroborated by the fact that an 'easy' display, such as that shown in Fig. 1, may be viewed with a variable-density filter in front of one eye. The effect will occur when this filter is set to pass so little light that the display coming to that eye is barely visible. Binocular difference-fields under these circumstances also must yield values which are non-zero.

(3) Length of shift. Thus far we have dealt only with lateral displacements of one column. Displacements of more than one column might be expected to enhance the separation of the two planes if the effect still occurred. It is reasonable to expect also that the effect would be more difficult to achieve with greater displacements. For an inner matrix of 32 columns and 32 rows, displays were generated with shifts of up to 9 columns. The stereoscopic effect was reported by some *O*s for all these displays, though some found it difficult to see the effect in the displays shifted 8 or 9 columns. The perceived separation of the two planes was indeed greater when the shift was longer. As Julesz has reported, there is a markedly increased latency of the stereoscopic effect with the more 'difficult' displays, some latencies being as long as 1 min. The *O*s initially report 'something different' about the inner area, but often it is some time before they report the spatial separation of the two planes.

The fact that the stereoscopic effect can occur with displacements of more than one column suggests that the visual system must in some sense slip the non-identical parts of the superimposed matrices past each other in search of a displacement which yields a better match. This displacement may be as large as 25% of the width of the inner matrix (9 columns in a 32-column matrix).

(4) Size, shape, and placement of the inner matrix. The displays thus far described have consisted of a square inner matrix, 32 × 32, centered in a square outer matrix, 64 × 64. Will the effect be manifested when the inner matrix is made very small or very large in relation to the outer matrix? In the limiting case, where the inner matrix was so enlarged as to make up the whole display, *i.e.* when it was 64 × 64, there was no effect with a one-column displacement. With an inner matrix 62 × 62 surrounded by an outer matrix only one row and one column wide around the periphery, the stereoscopic effect was, however, easily seen and was unmistakable. Under this condition, the effect was readily observed when any two sides were masked, or any three sides, although in the latter case the effect was considerably diminished.

An inner matrix consisting of a centered horizontal or a vertical band running the full width or the full length of the display also gives rise to the effect, even when the horizontal band has a thickness of only one row and when the vertical band is created by the interchanging of two adjacent columns. With the single horizontal row, several displays were generated to see how short this row could be made before the stereoscopic effect disappeared. Rows only five dots in length, lying in the central region of a 64 × 64 matrix, still gave rise to the stereoscopic effect, although *O* usually had to be told the exact region of the display in which to look for it. It is quite possible that the interchanging of only two adjacent dots in one row could give rise to the effect under ideal conditions, although this phenomenon has not been demonstrated.

(5) Matrix-density. In a study by Green, Wolf, and White of the detection of statistically defined patterns in dot-matrices, it was found that two methods of manipulating the grain of the display did not have similar effects on detection-thresholds.[2] Keeping the number of dots in the display constant and allowing area to vary with dot-departition was not equivalent to keeping the area of the display constant and allowing number of dots to vary with dot-separation. The authors suggested that the important parameter might be the ratio of dot-size to dot-separation.

Since Julesz speculated that the stimulus-parameters governing this stereoscopic effect might be the same as those governing the detection of contours in patterns such as those employed by Green, Wolf, and White, it seemed advisable to study them in this binocular display. With display-area constant and variation made in dot-separation together with number of dots, it was found that the stereoscopic effect became more difficult to achieve as dot-separation increased. The displays were approximately 1⅝ in. square and dot-size was constant. An arbitrary unit used in the computer-program for generating the displays is most convenient to use as a measure of dot-separation. In terms of this unit, two dots are approximately tangent when their dot-separation is 4. Dot-separation for this series was varied over a large range: from 5 to 257. At smaller dot-separations, 25 and below, the effect was easily achieved, but at larger separations, 51 and above, it was not.

The display with a dot-separation of 51—the densest display which failed to yield the stereoscopic effect—was photographically reduced until the dot-separation was below that of the display with a dot-separation of 25, which had yielded the effect. Even with this reduction, the effect could not be observed. Similarly, photographic enlargement of the 25 dot-separation display did not destroy the effect, thus lending support to the hypothesis that the ratio of dot-size to dot-separation is the important parameter.

The other approach to display-density left the number of dot-positions in the matrix fixed and allowed dot-separation and area to vary together. Since these displays were photographed with a camera different from that used in the previous series, it was to be expected that the dot-separation at which the effect changed would not be exactly comparable to that found earlier. In these displays a 16 × 16 matrix was used, the inner region being 8 × 8 and centered. At dot-separations of 12 and 16, the effect was clear, but with a dot-separation of 20 it was not.

---

[2] B. F. Green, Jr., A. K. Wolf, and B. W. White, The detection of statistically defined patterns in a matrix of dots, this JOURNAL, 72, 1959, 503-520.

(6) Shifting every Nth column. In the displays thus far discussed, every column in the inner matrix was shifted. Since Julesz suggests that the connectedness of the inner and outer regions is important in creating the effect, it is of some interest to examine displays in which some columns in the inner region are unchanged, while others are shifted. Three such sets were generated in which every second, every fourth, and every sixth column was shifted in the inner region, other columns remaining unchanged, and thus identical in the right and left displays. When every other column is shifted, the *O*s report a transparent screen of dots through which they can see other dots in the plane of the outer region. 'A beaded screen' is an apt description of the effect. This screen has sharp edges on four sides which give it definite shape. In the other two sets, with every fourth and every sixth column shifted, no such well-defined effect arises. Scattered dots and clusters of dots stand out from the plane of the outer region, but they do not form a bounded figure like the screen reported above.

(7) Non-random brightness-values in inner region. The brightness-values used in the displays considered thus far were randomly distributed over the inner and outer regions of the matrices with equal probability. Under these conditions, when the inner region has had its columns shifted one position, the distribution of brightness-differences between corresponding positions in the inner matrix has a constant expected value for all transpositions but one, the one which yields a large number of low or zero-differences. Were the brightness-values in the inner region differently distributed, this would not be the case. For example, if all the dots in the first column of the inner matrix had a brightness-value of 1, and if in each succeeding column the brightness-values were increased by 1, there would be a continuous brightness-gradient from left to right in the inner region. Under this condition, with a shift of one column position, the expected value of the distribution of brightness-differences also would vary continuously with the number of columns transposed. It would not have the sharp drop to a minimum characteristic of the random condition. This less clearly defined minimal difference might interfere with the occurrence of the effect. Nevertheless, displays generated with this regular brightness-increment in the inner region showed the stereoscopic effect clearly, although, because of the non-linearity of the brightness-increments, the inner region appeared curved rather than flat.

(8) Non-identical outer regions. The displays described thus far all have had outer regions which were identical for the two eyes. From some of the displays used by Julesz, it is clear that the corresponding outer regions do not have to be exactly the same, since he showed that the effect can occur even when one display is 10% larger than the other, and when Gaussian noise is introduced into one of the displays.

A more extreme difference between the outer region of the right and left displays would be obtained if the random distributions of brightness-values for the two were completely independent samples drawn from the same population. A pair of such displays was generated, in all other respects the same as the 'square within square' displays which had yielded a clear stereoscopic effect. The displays with the independent, random brightness-distributions in the outer regions did not yield a clear stereoscopic effect. Some *O*s had difficulty in defining the shape of the inner region which at first appeared round, but with no sharply defined contour between the inner and outer regions. One *O* reported a separation in depth between the two regions; others saw no such separation.

From the standpoint of the difference-field model, although there is a lateral displacement which will yield a high proportion of zero-differences in the inner region, the proportion of zero-differences in the outer region has an expected value of 25% when there are two brightness-values and no displacement, a value which will decrease slowly with increasing lateral displacement. These zero-differences, moreover, are scattered randomly throughout the region, and their loci also will change randomly with each lateral displacement. In this situation, therefore, there is no sharp peaking of zero-differences in the outer region, even though such a peak occurs in the inner region. Since the stereoscopic effect is absent or marginal under this condition, it may be that the effect requires some fairly sharp peaking of zero or at least small differences in both regions of the display.

(9) Monocular temporal integration. Although this effect is truly stereoscopic, it is conceivable that it might be achieved monocularly if the right and left matrices were presented to one eye in rapid succession. Should depth be seen under this condition, it would lend support to Julesz' contention that, after the difference-field has been obtained, the process of contour-formation is identical with that found in monocular perception.

This possibility was investigated by alternating the right and left matrices of a display such as that shown in Fig. 1 on 16-mm. motion picture film projected at varying speeds. At very slow rates of alternation, about 1 frame per sec., the dot-matrix appears to be stretching to the right and left in the central region. The boundaries of this stretch are not well defined, nor is there any perceptible difference in depth between the inner and outer regions. At slightly faster rates of alternation, the boundaries of the inner region become well defined and it appears to be sliding back and forth across the outer region, directly in front of it. At the top speed of this projector, about 24 frames per sec., the two regions become more distinctly separated in depth, though the separation is perhaps not as great as that reported under the stationary stereoscopic binocular viewing condition.

Discussion. This stereoscopic effect, which shows such an orderly relationship to display-parameters, offers a striking example of a perceptual phenomenon which must depend upon elaborate and precise processing of the images to the two eyes. Psychologists have been hesitant about postulating perceptual mechanisms which involve such data-processing of the sensory inputs, and have preferred to make the more or less gratuitous assumption that the perception is based on learning. The appeal to learning often is made to account for differing recognitive thresholds for various figures presented under identical conditions, or for certain visual illusions such as the rotating trapezoid of Ames.

With these stereoscopic displays, however, it is hard to see how a convincing case can be made for past experience in producing the effect. *O* need not be told either the location or the shape of the inner region beforehand, or even that he will see anything three-dimensional. The difference-field model offered by Julesz seems more appropriate, though it is admittedly a first approximation and in many respects incomplete. It specifies no boundary-values, for instance, which would enable one to predict threshold-values for the effect. As Julesz points out, an operation of either substraction or division would be possible in the light of the results thus far obtained, and the observations reported here are of little help in deciding which operation is the more reasonable. More precise control of dot-intensity in generating the displays, and actual photometric measurements of the displays, probably will be required to arrive at a decision.

One problem not dealt with in this model is one which has long been discussed: How are 'corresponding' regions in the right and left images identified? Historically, the concept of corresponding retinal points and the horopter has been invoked. Such an anatomically based concept has difficulty, however, in dealing with the fact that the stereoscopic effect can be observed when one image is 10% larger than the other in both height and width. The correspondence must be the result of prior processing of the images. The stereoscopic effect can be seen when the two differ not only in size but in the number of rows and columns. In such a case, how can the two images be so superimposed that a given dot in one matrix is fused with a given dot in the other matrix? Again it is plausible to postulate an operation which seeks some orientation of the two images' yielding a minimal sum of differences. Certain transformations on the images evidently are possible in this search: translation of one image relative to the other on the *X*-axis, and translation on the *Z*-axis. Furthermore, these transformations evidently need not be performed on the entire image, but may be done only on one or more circumscribed regions, in which case the fused binocular percept is split into regions separated in depth.

Many questions remain to be answered about this new and unusual stereoscopic effect. Can it be achieved if the elements differ in some other dimension than brightness—in size, for example, or in shape, or in color? For some time it has been known that the visual system responds to a disparity in size of the images to the two eyes with peculiar perceptions of curvature and tilt.[3] The response to an analogous disparity in over-all intensity of the images to the two eyes also is one of spatial displacement, under certain conditions.[4] Recently there has been considerable controversy over Land's demonstrations of the surprising response of the visual system when colors of the images to the two eyes differ in certain ways.[5] Here the binocular response differs from its monocular ingredients not in spatial respects, but in the distribution of perceived colors. In all of these cases, the response of the visual system to discrepant information at the two eyes appears to be a stable and orderly affair, under rather precise stimulus-control. Together they afford an unique opportunity to arrive at some understanding of the ways in which the incoming information is processed in the achievement of a unitary percept.

---

[3] Adelbert Ames, Binocular vision as affected by relations between uniocular

---

stimulus-patterns in commonplace environments, this JOURNAL, 59, 1946, 333-357; K. N. Ogle and P. Boeder, Distortion of stereoscopic spatial localization, *J. opt. Soc. Amer.*, 38, 1948, 723-733.

[4] Alfred Lit, The magnitude of the Pulfrich stereo phenomenon as a function of binocular differences of intensity at various levels of illumination, this JOURNAL, 62, 1949, 159-181.

[5] E. H. Land, Color vision and the natural image, *Proc. Nat. Acad. Sci.*, Parts I and II, 45, 1959, 115-129, 636-644.

CLASSICAL CONDITIONING IN THE FISH: FURTHER STUDIES OF PARTIAL REINFORCEMENT

By R. C. GONZALEZ, SANDRA MILSTEIN, and M. E. BITTERMAN,
Bryn Mawr College

In a recent study of classical conditioning in the goldfish, the effects of partial and consistent reinforcement on resistance to extinction were compared.[1] No significant difference in initial resistance was found in any of several comparisons involving various amounts of training, nor did a difference develop when partially and consistently reinforced animals were subjected to a series of extinctions separated by reconditioning sessions. The experiments to be reported here extend the range of circumstances under which the two kinds of reinforcement are compared.

EXPERIMENT I

Practice was relatively massed in previous work: 10 trials per day were given, with a mean intertrial interval of 3 min. In the present case, by contrast, the intertrial interval was 24 hr. Interest in spaced practice arose out of the results of experiments with another fish—the African mouthbreeder—in a simple instrumental situation.[2] In those experiments, initial resistance to extinction was significantly greater after consistent than after partial reinforcement both with massed and with spaced practice, but the difference was much more striking with spaced practice.

Method. The *Ss* were 36 experimentally naïve goldfish, about 2 in. long, obtained from a local dealer. They were maintained in individual 2-gal. tanks set on open shelves in the laboratory.

The apparatus employed is described in detail elsewhere.[3] It consists of six conditioning chambers and an amplifying-integrating device for measuring general ac-

---

* Received for publication January 7, 1961. This work was supported by Grant M-2857 from the National Institute of Mental Health.

[1] R. C. Gonzalez, Nicholas Longo, and M. E. Bitterman, Classical conditioning in the fish: Exploratory studies of partial reinforcement, *J. comp. physiol. Psychol.*, 54, 1961, 452-460.

[2] Jerome Wodlinsky and M. E. Bitterman, Partial reinforcement in the fish, this JOURNAL, 72, 1959, 184-199; Resistance to extinction in the fish after extensive training with partial reinforcement, this JOURNAL, 73, 1960, 429-434; Nicholas Longo and M. E. Bitterman, The effect of partial reinforcement with spaced practice on resistance to extinction in the fish, *J. comp. physiol. Psychol.*, 53, 1960, 169-172.

[3] J. L. Horner, Nicholas Longo, and M. E. Bitterman, A classical conditioning technique for small aquatic animals, this JOURNAL, 73, 1960, 623-626.

tivity. Each chamber is a covered 2-gal. aquarium, painted black, containing a small screened enclosure into which *S* is placed. Just outside the enclosure is a thin metal paddle which is moved by currents of water produced by *S*'s activity. The paddle is fixed to a light rod, the upper end of which is inserted into the needle-holder of a crystal phonograph cartridge. The cartridge is mounted on the cover of the tank, and its output is led to the amplifying-integrating circuit. The activity which occurs in any of the conditioning chambers during a given period of time may be read by *E* from a meter in the control panel. The conditioning chamber in which activity is to be measured is determined by a selector switch, and the interval of time during which activity is measured is set on a Hunter timer. Suspended about 1 in. above the center of *S*'s enclosure is an inside-frosted 10-w. lamp. This lamp is the only source of illumination when the chamber is covered, and either its onset or its offset—in this experiment, its onset—serves as the *CS*. All stimuli are presented automatically after *E* presses a button to initiate a trial.

The fish were trained in groups of 6 *S*s each. They were carried individually from home tank to conditioning chamber by means of a net, with *E*'s hand gently closed around the fish's body, and placement into *S*'s enclosure was accomplished by hand. Following transfer, a few minutes were allowed for adaptation to the apparatus, during which time *E* cleaned the home tanks and restored the standard water level.

Throughout the experiment, one trial was given each day. In the first stage, which lasted 15 days, each trial was reinforced for all *S*s. The *CS–US* interval was 4.5 sec., and the termination of the *CS* coincided with the termination of the *US*, a shock of 0.5-sec. duration. Response to the *CS* was measured during the 4.5-sec. period of stimulation prior to the onset of shock. On the basis of their performance during the first stage of training, the *S*s were divided into two matched groups—a Consistent Group and a Partial Group.

In the second stage of the experiment, which lasted 46 days, each trial was reinforced for the *S*s of the Consistent Group, while only 23 of the trials were reinforced for the *S*s of the Partial Group, the sequence of reinforced and unreinforced trials following the rules of Gellermann.[4]

In the third stage, all *S*s were extinguished. On each daily trial, the *CS* was presented alone for 5 sec., and the response to the *CS* was measured during the first 4.5 sec. of stimulation. A criterion of extinction—response-magnitude of 1.0 or less on three successive days—was adopted in advance, but work with each *S* continued until all except a very few had met the criterion.

Two measures of basal activity were made each day during the 4.5-sec. periods of no stimulation, one shortly before and one shortly after the trial. After the second activity-measure, the *S*s were returned to their living tanks, and another set of *S*s was introduced into the conditioning chambers.

Results. In Fig. 1, the main results of the experiment are summarized in terms of mean magnitude of response to the *CS*. The first 25 days of training are represented, as are the 28 days of extinction. The intervening 36 days of training, during which the groups continue to respond at the levels attained earlier, are omitted in the interest of space.

---

[4] L. W. Gellermann, Chance orders of alternating stimuli in visual discrimination experiments, *J. genet, Psychol.*, 42, 1933, 206-208.

In acquisition, both before and after the introduction of partial reinforcement, the performance of the two groups was very much the same. A Lindquist Type I analysis shows the effect of Days to be significant beyond the 5% level—the level adopted in all tests reported here—($F = 10.50$ with 60 and 1960 *df*); the effect of Conditions does not approach significance, and the interaction is negligible. The analysis is based on the data of only 33 *S*s (17 in the Consistent Group and 16 in the Partial Group), because three animals were lost during this phase of the work.

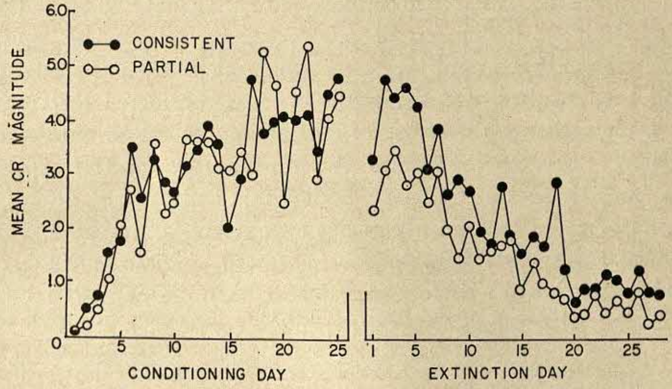


FIG. 1. MEAN MAGNITUDE OF RESPONSE TO THE *CS* DURING THE FIRST 25 DAYS OF CONDITIONING AND DURING EXTINCTION IN EXPERIMENT I

In extinction, there was a slight tendency for the Consistent Group to respond at a higher level, but the difference is not significant ($F = 2.63$ with 1 and 30 *df*). Only the effect of Days is significant ($F = 12.75$ with 27 and 810 *df*). This analysis is based only on the data of 32 *S*s (16 in each group), because an additional animal was lost in extinction.

The experiment was terminated after 28 days of extinction, at which time all but four animals (two in each group) had met the criterion. It did not seem reasonable to continue longer, because the response-magnitudes of the four exceptional *S*s seemed to be near asymptotes somewhat above 1.0 (the criterion value). Nevertheless, resistance to extinction in the two groups may be compared in terms of days to criterion. If the four animals which did not reach the criterion are assigned scores of 31 (the minimal number of days in which they might have reached the criterion had the experiment been continued), the means are 17.6 days for the Consistent Group and 14.7 days for the Partial Group. By Wilcoxon's nonparametric method for paired replicates, the difference is not significant.

Mean basal activity remained low (less than 1.0) for both groups throughout the experiment, the difference between the two groups in this respect being negligible.

EXPERIMENT II

Experiment II was concerned with repeated conditioning and extinction: conditioning days were alternated with extinction-days for partially and for consistently reinforced animals. One purpose of this experiment was to compare the resistance to extinction of Partial and Consistent Groups after relatively small numbers of reinforcements. In previous work, extinction was not begun until long after asymptotic levels of response to the *CS* had been reached, and the possibility was considered that measurements made earlier in the course of conditioning might be more sensitive to differences between the groups. A second purpose of this experiment was to study changes in resistance, both within and between groups, over the series of extinctions. In a previous experiment with a food-rewarded instrumental response, consistently reinforced African mouthbreeders showed greater initial resistance to extinction than did partially reinforced controls, but relative resistance shifted in the course of a series of extinctions separated by retraining sessions, the absolute resistance of the Consistent Group declining more rapidly than that of the Partial Group.[5] In the rat, too, resistance to extinction declines progressively under such conditions, more rapidly with consistent than with partial reinforcement, although resistance after partial reinforcement is, of course, greater to begin with.[6] No indication of a change in relative resistance appeared in the first work on classical conditioning in the goldfish, but the number of extinctions studied was not large.[7]

Method. The *S*s were 42 experimentally naïve goldfish, drawn from the same population and maintained in the same manner as those of Experiment I. Prior to the beginning of training, the animals were divided at random into two groups, Partial and Consistent.

The apparatus was the same as that used in Experiment I. Ten trials per day were given, with a systematically varied intertrial interval averaging 3 min. Again the *CS* was light-onset. The *CS–US* interval was 4.5 sec., and the termination of the *CS* coincided with the termination of the *US*, a shock of 0.5-sec. duration. Response to the *CS* was measured during the first 4.5 sec. of the *CS*. Shortly before and shortly after each daily series of trials, basal activity was measured during 4.5-sec. periods of no stimulation.

---

[5] Wodinsky and Bitterman, *op. cit.*, this JOURNAL, 72, 1959, 189-190

[6] C. C. Perkins, Jr., and A. J. Cacioppo, The effect of intermittent reinforcement on the change in extinction rate following successive reconditionings, *J. exp. Psychol.*, 40, 1950, 794-801.

[7] Gonzalez, Longo, and Bitterman, *op. cit.*, 452-460.

The duration of the experiment was 26 days. On odd days, the animals were conditioned. The Consistent Group was given 10 reinforced trials; the Partial Group was given 5 reinforced and 5 unreinforced trials in Gellermann orders, the first trial being reinforced on some days and unreinforced on others. On even days, the animals were extinguished, both groups being given 10 unreinforced trials.

Results. Mean magnitude of response to the *CS* on conditioning days is plotted in the left-hand portion of Fig. 2. The two curves have much the same shape, rising sharply and then declining to what seems to be a stable level, but the magnitude of response is substantially higher in the Consistent Group. The variance due to Conditions is significant ($F = 4.31$, $df = 1, 31$), as is the variance due to Days ($F = 9.58$, $df = 12, 372$), but that due to interaction is not ($F = 1.09$, $df = 12, 372$). The analysis is based on the data for 17 Consistent and 16 Partial *Ss*, several *Ss* having been lost in the course of the experiment.

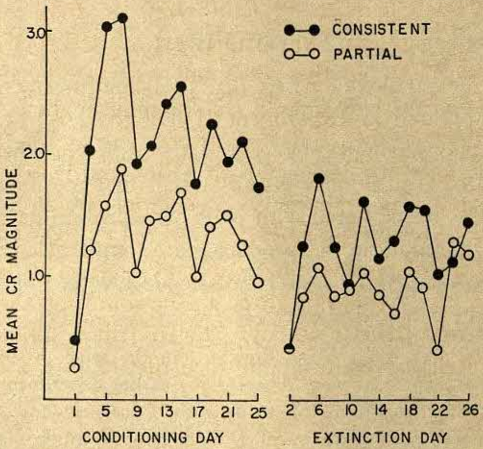


FIG. 2. MEAN MAGNITUDE OF RESPONSE TO THE *CS* ON CONDITIONING AND ON EXTINCTION-DAYS IN EXPERIMENT II

Mean magnitude of response to the *CS* on extinction-days is plotted in the right-hand portion of Fig. 2. The curve for the Consistent Group soon reaches a higher level (indicating greater resistance to extinction) than that of the Partial Group, and there is no appreciable change in the level of either group after the third extinction. The curves cross on Day 24, but that is due to a rise (probably spurious) in the level of the Partial Group rather than to a more rapid decline in the level of the Consistent Group. In fact, neither group gives any real indication of a progressive decline in resistance to extinction in the course of the experiment, although analysis of variance shows the effect of Days to be significant ($F = 5.92$ with 12 and 372 *df*). The effect of Conditions is not significant ($F = 1.90$ with 1 and 31 *df*), nor is the interaction ($F = 1.51$ with 12 and 372 *df*).

The pattern of change in *CR* magnitude which is shown in Fig. 2 is not reflected in the basal activity measures. For both groups, basal activity was of very low magnitude, well below 1.0. Responses of greater magnitude appeared only infrequently, were not characteristic of any individual *S* or group, and showed no change in frequency over the training series.

EXPERIMENT III

No sign of the paradoxical effect of partial reinforcement on resistance to extinction appeared in Experiment II, but the level of acquisition in the Partial Group was rather low, and the difference between the two groups in level of acquisition was substantial. In Experiment III, pairs of conditioning days were altered with pairs of extinction-days. It was anticipated that the procedure would produce more comparable levels of acquisition in the two groups and therefore provide greater opportunity for the appearance of the paradoxical effect.

Method. The Ss were 60 goldfish drawn from the same population and maintained in the same manner as were those of the earlier experiments. They were assigned at random to two groups of 30 *Ss* each, Consistent and Partial. The procedure was the same as in Experiment II with two exceptions: (1) two days of conditioning were alternated with two days of extinction for a total of 60 days; (2) for all animals, each conditioning day began with a reinforced trial. The amplifier-integrator used in this experiment was modified somewhat in the interest of greater linearity, and the absolute values of the readings obtained became somewhat lower in the process.

Results. Mean magnitude of response to the *CS* on the first of each pair of conditioning days is plotted for each of the two groups in the left-hand portion of Fig. 3, and the corresponding values for the first extinction day of each pair are plotted in the right-hand portion. (The curves for the second conditioning and extinction-days are not plotted because of their close resemblance to the ones presented.) Clearly, the difference between groups on conditioning days is much smaller in this experiment than in Experiment II, although it is statistically significant ($F = 5.34$ with 1 and 41 *df*).[8] The effect of Series is significant ($F = 55.57$ with 14 and 574

---

[8] The analysis is based on the data for 24 *Ss* in the Consistent Group and 19 in the Partial Group. A substantial group of *Ss* was lost due to equipment failure, and a few others died in the course of the experiment.

df), as is the interaction ($F = 3.68$ with 14 and 574 *df*). On extinction-days, the difference between groups disappears altogether. As the experiment continues, resistance to extinction increases in both groups, and no sign of the paradoxical effect appears.

Mean within-days curves are plotted in Fig 4. They are based on the last 10 pairs of conditioning and extinction-days, excluding the early stages of the experiment in which changes between sessions were most rapid. The extinction-curves show a good deal of within-days decline in level of response, but the conditioning curves show little in the way of growth, except perhaps for the Day 1 curve of the Consistent Group. Spontaneous recovery from the last trial of extinction to the first trial of extinction, as well as from the tenth to the eleventh trial of extinction, is very high.

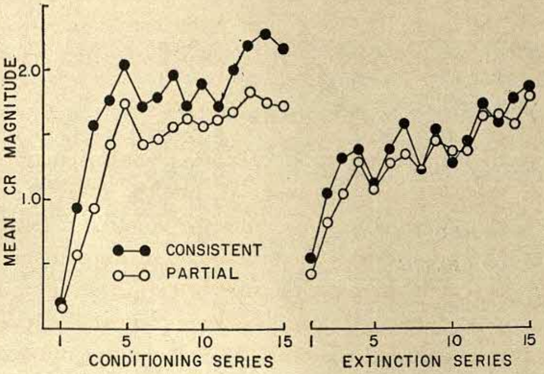


FIG. 3. MEAN MAGNITUDE OF RESPONSE ON THE FIRST OF EACH PAIR OF CONDITIONING DAYS AND ON THE FIRST OF EACH PAIR OF EXTINCTION-DAYS IN EXPERIMENT III

Throughout the experiment, intertrial activity remained low. Systematic measurements made during periods of no stimulation did not reflect changes in magnitude of response to the *CS*.

SUMMARY AND CONCLUSIONS

Three experiments on classical conditioning in the goldfish were performed to study the effect on resistance to extinction of partial as compared with consistent reinforcement. In Experiment I, practice was spaced (the intertrial interval was 24 hr.). In the remaining experiments, with massed practice, repeated conditioning and extinction was studied; conditioning and extinction-days were alternated in Experiment II, while pairs of conditioning and extinction-days were alternated in Experiment III.

Both in these experiments and in a set reported in an earlier paper, there was a tendency—sometimes significant, sometimes not—for level of acquisition to be lower after partial than after consistent reinforcement. In none of the experiments was there any indication of the paradoxical effect of partial reinforcement on initial resistance to extinction which appears so dependably in work with mammals. Nor was there any change in relative resistance over a series of repeated extinctions of the kind shown both by the rat and by the African mouthbreeder in instrumental situations; there was, in fact, no clear indication at all of progressive decline in resistance to extinction under either condition of reinforcement.

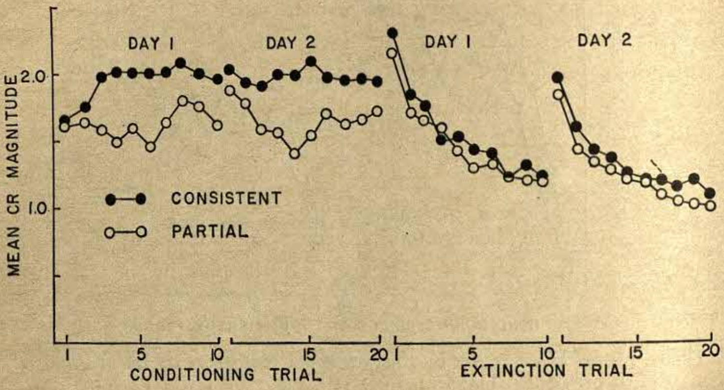


FIG. 4. WITHIN-DAYS CONDITIONING AND EXTINCTION-CURVES FOR EXPERIMENT III

Further experiments are necessary to separate the various factors which may be responsible for these differences. One factor is species; it was, of course, an interest in species-differences which gave rise to the research program of which the present work is a part.[9] Another factor is method; except for some experiments with human *S*s, almost nothing has been done until now on the problem of resistance to extinction following partial reinforcement in classical conditioning. All that can be said at present is that the generality of the paradoxical effect and related phenomena seems markedly to have been overestimated.

---

[9] Bitterman, Toward a comparative psychology of learning, *Amer. Psychologist,* 15, 1960, 704-712.

S–R AND R–S LEARNING AS FUNCTIONS OF INTRALIST SIMILARITY

By SLATER E. NEWMAN, North Carolina State College, and
ROBERT BUCKHOUT, Wright Air Development Center

Research in verbal learning has demonstrated that during the learning of S–R pairs, *S* may be learning R–S pairs as well.[1] Perhaps the most parsimonious view of R–S learning is that it is governed by the same laws which govern S–R learning. If this proposition is correct, then those variables which affect the rate at which S–R pairs are learned can be expected to affect similarly the rate at which R–S pairs are learned.

Previous studies of S–R learning have demonstrated that an increase in intralist similarity has generally been accompanied by a reduction in rate of learning the S–R list.[2] Thus, in the present study it was anticipated

---

* Received for publication December 14, 1959. Based on an M.S. thesis at North Carolina State College by the junior author. This report was prepared by the senior author. Preparation was in part supported by Office of Naval Research Contract Nonr 486 (08). This is Technical Report Number 2 of the series prepared under this contract.

[1] Hermann Ebbinghaus, *Ueber das Gedachtnis: Untersuchungen zur experimentellen Psychologie,* 1885, 1-169; (Translated as *Memory: A Contribution to Experimental Psychology,* by H. A. Ruger and C. E. Bussenius, 1913); A. Wohlgemuth, On memory and the direction of associations, *Brit. J. Psychol.,* 5, 1913, 447-465; H. E. Garrett and G. W. Hartman, An experiment on backward association in learning, this JOURNAL, 37, 1926, 241-246; T. G. Hermans, A study of the relative amounts of forward and backward associations of verbal material, *J. exp. Psychol.,* 19, 1936, 769-775; J. A. McGeoch, The direction and extent of intraserial associations at recall, this JOURNAL, 48, 1936, 221-245; J. T. Wilson, Remote associations as a function of the length of interval between learning and recall, *J. exp. Psychol.,* 33, 1943, 40-49; The formation and retention of remote associations in rote learning, *ibid.,* 39, 1949, 830-838; E. R. Harcum, Verbal transfer of overlearned forward and backward associations, this JOURNAL, 66, 1953, 622-625; Yasuo Morikawa, Studies in paired associate learning: I. Forward-backward recall gradient, *Jap. J. Psychol.,* 26, 1955, 156-171; B. B. Murdock, Jr., "Backward" learning in paired-associates, *J. exp. Psychol.,* 51, 1956, 213-215; "Backward" association in transfer and learning, *ibid.,* 55, 1958, 111-114; Samuel Feldman and B. J. Underwood, Stimulus recall following paired-associate learning, *ibid.,* 53, 1957, 11-15; E. M. Jantz and B. J. Underwood, R–S learning as a function of meaningfulness and degree of S–R learning, *ibid.,* 56, 1958, 174-179; R. G. Hunt, Meaningfulness and articulation of stimulus and response in paired-associate learning and stimulus recall, *ibid.,* 57, 1959, 262-267.

[2] E. J. Gibson, Intra-list generalization as a factor in verbal learning, *ibid.,* 30, 1942, 185-200; Underwood, Studies of distributed practice: X. The influence of intralist similarity on learning and retention of serial adjective lists, *ibid.,* 45, 1953, 253-259; B. J. Underwood and Jack Richardson, The influence of meaningfulness, intralist similarity, and serial position on retention, *ibid.,* 52, 1956, 119-126.

that an increase in intralist similarity, whether among stimulus-terms or among response-terms, would lead to fewer elicitations of the S-term when an R-term is used as the stimulus following S–R learning. The effect of increased similarity would, according to generalization-theory, be expected to lead to an increased tendency for generalization to occur and a concomitant decrease in the rate at which R–S associations would be established. (Since Gibson's theory deals only with the effects of variation in similarity among stimulus-terms, extension to deal with the effects of variation in similarity among response-terms as well, is referred to in this paper as 'generalization-theory').[3]

Data bearing on these hypotheses have been reported by Feldman and Underwood.[4] They found, in keeping with generalization-theory, that as similarity among stimulus-terms of an S–R list was increased, the number of correct elicitations of the S-term when the R-term was the stimulus decreased. Increase in similarity among response-terms did not, however, lead to the anticipated decrease in number of correct elicitations of the S-term during R–S recall. The effect on S–R learning of variations in similarity among both S-terms and R-terms occurred as anticipated. Thus, only in the case of the effects of the similarity among R-terms on the measurement of R–S retention, did the data fail to support predictions based on the generalization-theory.

Study of the procedures employed by Feldman and Underwood suggested two factors which might account for such a divergence. First, the S-terms were nonsense-syllables low in associative value, while the R-terms were meaningful adjectives. Since meaningful terms are more highly differentiated than meaningless terms, it seemed likely that the failure to find a difference in R–S recall as a function of R-term similarity might have resulted from the use of meaningful terms in the response-position. Use of a criterion of two perfect trials also may have contributed to the failure of similarity among R-terms to affect R–S recall. The use of such a criterion presumably could have made more equal the associative strengths for R–S pairs among the experimental lists, reducing the likelihood that the manipulated variables would be as effective. Data on the effects of these factors have been presented by Jantz and Underwood and by Hunt.[5] Number of R–S associations was found to be directly related to degree of S–R learning by Jantz and Underwood and to R-term meaningfulness ($0.10 > p > 0.05$) by Hunt.

The present experiment, as did the Feldman and Underwood study, sought to investigate the effects of variation in similarity among S-terms and among R-terms on R–S recall following S–R learning, as well as their effects on rate of S–R learning. The Feldman and Underwood procedure was modified, however, in two respects:

---

[3] Gibson, A systematic application of the concepts of generalization and differentiation to verbal learning, *Psychol. Rev.* 47, 1940, 196-229.

[4] Feldman and Underwood, *op. cit.,* 11-15.

[5] Jantz and Underwood, *op. cit.,* 174-179; Hunt, *op. cit.,* 262-267.

(1) For both S-terms and R-terms, nonsense-syllables of low associative value were employed, and

(2) Learning was continued for a fixed number of trials rather than to a criterion defined by performance.

In the present study, the following four predictions were tested:

Prediction I. As similarity among the stimulus-terms of S–R pairs is increased, the number of S-terms correctly elicited using the R-terms as a stimulus following S–R learning will decrease.

Prediction II. As similarity among the response-terms of S–R pairs is increased, the number of S-terms correctly elicited using the R-term as a stimulus following S–R learning will decrease.

TABLE I

LISTS FOR EACH EXPERIMENTAL GROUP

LL	*LH*	*HL*	*HH*
TOV-GAH	TOV-ZAJ	VAF-SIJ	VAF-ZIQ
VUF-QOJ	VUF-QOJ	VUX-QOJ	VUX-QOJ
XIF-JUH	XIF-JUH	XIF-JUH	XIF-JUH
MEF-SIJ	MEF-QIH	XAV-CEJ	XAV-QIH
NUX-LAJ	NUX-QIJ	XUY-GAH	XUY-ZAJ
YUV-ZEQ	YUV-ZEQ	YUV-ZEQ	YUV-ZEQ
DAX-CEJ	DAX-ZIQ	YOV-LAJ	YOV-QIJ

Prediction III. As the degree of similarity among stimulus-terms of S–R pairs is increased, the rate of S–R learning will decrease.

Prediction IV. As the degree of similarity among response-terms of S–R pairs is increased, the rate of S–R learning will decrease.

METHOD

Subjects. A group of 80 men volunteers, enrolled as undergraduates at North Carolina State College, served as *S*s. They were assigned to four experimental groups of 20 each on a systematic basis. Only the differences in lists learned differentiated the treatment of each group.

Materials. Four lists varying in degree of similarity among stimulus- and response-terms were constructed. (See Table I) The designation and description of each list follow:

List *LL,* low similarity among both stimulus- and response-terms.
List *LH*, low similarity among stimulus-terms and high similarity among response-terms.
List *HL,* high similarity among stimulus-terms, low similarity among response-terms.
List *HH,* high similarity among both stimulus- and response-terms.

Each list was prepared in accordance with the following criteria:

(1) Nonsense-syllables using a *CVC* pattern were employed. (2) Variations in similarity were achieved through restricting the number of consonants used. For

the high stimulus-similarity lists, the four consonants FVXY were employed; for the low stimulus-similarity lists, the eight consonants, FVXYDMNT. For the high response-similarity terms, the four letters HJQZ formed the consonant pool; for the low response-similarity terms the eight letters HJQZCGLS were used. (3) In each list, both for stimulus- and for response-terms, there were four nonsense-syllables of 0% associative value from Glaze's lists and three syllables of 7% associative value.[6] (4) The same vowel did not appear in both the stimulus- and response-member of a pair. (5) An attempt was made to rule out predictable letter-sequences and obvious S–R associations between items of a pair.

Procedure. Initially, instructions were read to *S* describing the paired-associate procedure. *S* then learned a list of three paired associates of 100% associative value from Glaze's lists as training to a criterion of one perfect trial. The list appropriate to his experimental group was then exposed to each *S* for 39 trials. The lists were presented on a memory-drum at 6-sec. rate, 3 sec. for the stimulus and 3 sec. for stimulus and response. Three different orders of the syllables were used to reduce the likelihood that serial learning would occur. The items were typed in upper case letters, and were viewed through an opening in a masonite wall separating *E* and *S*. During learning, *E* recorded all responses made by *S*.

Immediately after Trial 39, each of the response-terms was presented for 2.5 sec. after which 2.5 sec. was allowed for *S* to spell out the stimulus-term that had been paired with it during S–R learning. The order during this test was different from the order employed during learning. Following this recall (the drum-recall), each response-term typed on a 3 × 5 card was presented in a different order and *S* was instructed to spell out the stimulus-term which had been associated with it during S–R learning. No time limit was imposed during this period of unlimited recall. A brief post-experimental inquiry followed for each *S* after which he was dismissed.

RESULTS

Three response-measures have been employed; the number of syllables correct, the number of letters correct (*i.e.* number of letters in correct position for a syllable), and the number of different responses from the list without reference to position (response-terms during S–R Trial 39, stimulus-terms during R–S recall). The data for each response measure under each condition are summarized in Table II. The significance of these results was evaluated by means of a 2 × 2 analysis of variance for each column of Table II. These analyses are summarized for R–S recall and for S–R Trial 39 below.[7]

An inspection of Table II makes clear that in every instance in the first six columns, the greater is the degree of similarity, the poorer is the recall of S-terms. This is true both for similarity among S-terms and for similarity among R-terms. It is true for recall when R-terms are presented on the drum for a brief period and when the R-terms are presented on cards with no limit on recall. It is true when number of syllables correct, letters correct, or different responses is the dependent variable.

---

[6] J. A. Glaze, The association value of nonsense-syllables, *J. genet. Psychol.*, 35, 1928, 255-269.

[7] Dr. C. W. Gray assisted the author in the statistical treatment of the data. All analyses of variance were based on distributions for which the variances were found to be not significantly heterogeneous, using Bartlett's test ($p > 0.01$).

The analyses of variance for these 6 columns show that of 12 tests covering the factor of similarity, 8 reach significance at a level better than $p = 0.05$; 4 of these are for stimulus-similarity, and four for response-similarity.

The results for correct responses during R–S recall thus tend to support Predictions I and II. There is some evidence, in the case of letter-scores under unlimited recall, to suggest that the effects of varying stimulus-similarity may differ at high and low levels of response-similarity. The results for different responses indicate that the number of different S-terms emitted during R–S recall is inversely related both to S-term and to R-term similarity.

TABLE II

Mean Number of Syllables Correct, Letters Correct and Different Responses. Groups Differ in Similarity of Stimulus and Response Syllables.

	Drum-recall			Unlimited recall			S–R Trial 39		
Group	syllables	letters	different responses	syllables	letters	different responses	syllables	letters	different responses
HH	.55	4.10	.85	1.25	7.65	2.05	2.65	11.05	3.35
HL	.85	5.55	1.30	1.65	10.05	2.30	3.70	12.50	4.25
LH	1.10	5.20	1.40	1.60	7.15	2.20	2.55	10.90	2.85
LL	2.00	8.45	2.20	3.00	12.75	3.20	5.60	17.55	5.75

S–R learning. Predictions III and IV were concerned with the effects of similarity on the learning of the S–R pairs. Learning curves for the four groups are shown in Fig. 1. They show a steady, almost linear, increase over the 39 repetitions of the list. As was predicted, the most rapid learning was for the *LL* group, exposed to the most differentiated stimulus- and response-elements. The remaining three groups fall rather close together, the *HL* group being slightly better than either the *LH* or *LL* groups.

Three analyses of variance were done on the results for Trial 39. The means are presented in Table II, and the analyses are given below.

(1) Number of syllables correct. The analysis of variance of these data gave an F for stimulus-similarity, for response-similarity and for the S × R interaction that are all significant. The cell means vary in the predicted direction except that at high response-similarity an increase in degree of similarity among the stimulus terms was accompanied by a slight increase in the number of correct S–R pairs. Use of Tukey's studentized range-test indicated that only the difference between the *LH* and *LL* means was significant ($p < 0.01$). The differences involving the *HH* vs. *LH*, *HH* vs. *HL* and *HL* vs. *LL* means were not found to be significant ($p > 0.05$).

(2) Number of letters correct. The analysis of variance yielded an F for stimulus-similarity of 5.86 ($p < 0.05$), for response-similarity of 16.03 ($p < 0.01$), and for interaction of 6.61 ($p < 0.05$). Examination of the cell means again shows that they vary in the predicted direction except that at high response-similarity, an increase in degree of similarity among stimulus-terms was accompanied by a slight increase in number of correct letters among S–R pairs. Use of Tukey's studentized range-test indicated that the difference involving the *LH* and *LL* means was significant ($p < 0.05$). The differences involving the *HH* vs. *HL, HH* vs. *LH,* and *HL* vs. *LL* means were not found to be significant ($p > 0.05$).

(3) Number of different responses. The analysis of variance gave significant *F*s ($p < 0.01$) for response-similarity and interaction and a nonsignificant F ($p > 0.05$) for stimulus-similarity.[8] Use of Tukey's studentized range-test indicated that the difference involving the *LH* and *LL* means was significant ($p < 0.05$). The differences involving the *HH* vs. *HL, HH* vs. *LH* and *LH* vs. *LL* means were not found to be significant ($p > 0.05$). Thus, increase in R-term similarity led to emission of fewer different R-terms during S–R Trial 39, particularly when stimulus-similarity was low.

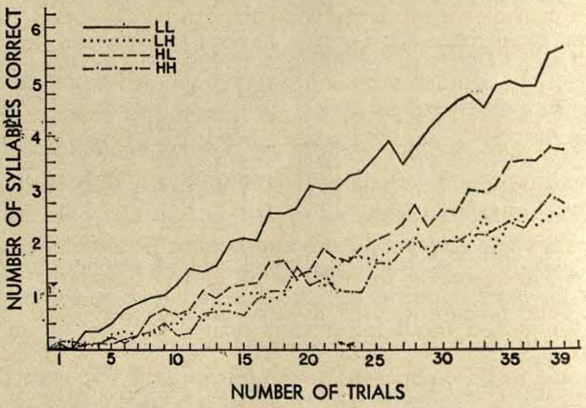


FIG. 1. LEARNING CURVES FOR EACH OF THE FOUR GROUPS

The results for correct responses did not completely support Predictions III and IV. As the degree of similarity among S-terms and among R-terms increased the rate of S–R learning tended to decrease, but at high levels of response-similarity, variation in stimulus-similarity, did not lead to the anticipated change in performance.

---

[8] J. O. Cook, Response analysis in paired-associate learning, in A. A. Lunsdaine (Ed.), *Student Response in Programmed Instruction,* in press.

DISCUSSION

The inverse relationships between number of correct responses during S–R Trial 39 and both S-term and R-term similarity tend to support the predictions derived from generalization-theory. The inverse relationship between R-term similarity and number of correct responses plus the findings of Horowitz and of Underwood, Runquist, and Schulz suggest that increased R-term similarity interfered with associative learning.[9] The decrease in number of different responses emitted as R-term similarity increased may also have been due to interference with associative learning rather than with response-learning. This is suggested by the results of an analysis of the data for the trial on which each R-term was initially emitted. High response-similarity was found to be associated (though not significantly) with early emission of the R-term.

Increase in the similarity among S-terms also led to fewer correct responses on S–R Trial 39. The inverse relationship observed between S-term similarity and number of different S-terms emitted during R–S recall suggests that S-term discrimination as well as S–R associative learning may have been interfered with.[10]

The findings for R–S recall also support predictions from generalization-theory. The decrease in number of correct responses during R–S recall as a function of increased R-term similarity may have been due to interference with R–S associative learning or, as the decrease in number of S-terms suggests, to interference with learning the responses (mediating and overt) for the S-terms.[11] The decrease in both number of correct responses and number of different responses during R–S recall as a function of increased S-term similarity may reflect interference with learning the responses for the S-terms, with R–S associative learning or with both. The tenability of the conclusions for S-term discrimination and learning depend on whether the number of different S-terms emitted during R–S recall is a valid indicator of the number of the stimulus terms to which the responses (mediating and overt) have been learned.[12]

---

[9] L. M. Horowitz, Associative matching and intralist similarity, personal communication, 1961; B. J. Underwood, W. N. Runquist, and R. W. Schulz, Response learning in paired-associate lists as a function of intralist similarity, *J. exp. Psychol.*, 58, 1959, 70-78.

[10] S. E. Newman, A mediation model of paired-associate learning, *Office of Naval Research Tech. Rep. No. 1,* Contract Nonr 486(08), 1961.

[11] Newman, *op. cit.,* 3.

[12] L. R. Taylor, The effects of change in context elements on performance following paired-associate learning as a function of element-sharing among stimulus-terms. Unpublished M. S. thesis, North Carolina State College, 1962.

SUMMARY

Results of the present study suggest that as similarity is increased among S-terms and among R-terms, the number of correct responses during R–S recall decreases and (at least at low levels of R-term similarity) the rate of S–R learning tends to decrease. These findings tend to support predictions based on generalization-theory. Possible loci of interference due to increase in S-term and R-term similarity were identified.

INDIVIDUAL DIFFERENCES IN THE STRUCTURE OF COLOR-PERCEPTION

By CARL E. HELM, Princeton University, and LEDYARD R. TUCKER,
University of Illinois

A problem which has intrigued both physicists and psychologists for many years is the determination of a metric for perceived color-space. A metric color-space is a representation of the relations among colors in a model in which the distance between a pair of colors gives their perceptual difference, and direction coincides with attribute. Helmholtz was the first to look for a solution, and a committee of the Optical Society of America is the latest, being actively engaged in work on the problem at the present. Judd discussed the situation as follows:

> A uniform tridimensional color scale would be an important commercial, as well as scientific, achievement. It would simplify color specification and the setting of color tolerances, it would clarify the interpretation of one-dimensional color scales applied to off-color samples, it would be an important guide to the preparation of reference color samples as standards, and it would aid selection of harmonious color combinations. Unfortunately, scores of attempts have so far not succeeded in developing such a scale.[1]

The ideal which has been sought for so many years now with only limited success may be more precisely described as a three-dimensional Euclidean space in which each of the dimensions correlates highly with one of the one-dimensional color scales (lightness, saturation, or hue) and in which distance correlates highly with the amount of perceived difference.

In recent years, there have been several investigations of the perceived relations among colors utilizing multidimensional scaling procedures. At least part of the motivation for these investigations came out of the conviction that these (possibly more appropriate) techniques would facilitate the development of a uniform color-scale.

During the course of some recent work at Princeton, a different possible source of difficulty became apparent. Research in this area generally is carried out with *Ss* having normal color vision, normal in terms of their being able to pass one or the other of the standard tests for color-blindness. We used the Hardy, Rand, and Rittler pseudoisochromatic charts, published by the American Optical Company, to sort our *Ss* into normals and color-deficients. It was possible to distinguish among the color-deficient *Ss* quite readily in terms of the differing degrees of difficulty they encountered in reading the charts. The normal *Ss* however, were generally indistin-

---

* Received for publication January 18, 1961.

[1] D. B. Judd, *Color in Business, Science, and Industry*, 1952, 224.

guishable from one another in terms of their performance. In terms of the language of scaling, these color-blindness tests involve something less than ordinal measurement, since *S* is able to carry out the task if he perceives two colored regions as being distinct in a particular way, and he fails the task if he fails to perceive this difference. We might be able to locate our *S*s on an ordinal scale or perhaps a crude sort of interval-scale if we questioned them closely in regard to whether or not the two regions were decidedly different or only just barely different. The possible source of difficulty arises when, in carrying out color-scaling experiments, we go on to require these normal *S*s to make ratio-judgments of a more or less complex and precise nature and the results are combined across *S*s. Quite possibly, the *S*s who do not differ on ordinal-scaling will differ in making ratio-judgments, and to combine data might be to obscure genuine perceptual differences. The investigation of these perceptual differences is made doubly important in view of the fact that this kind of research is generally carried out with small groups of *S*s.

In our research, we had the *S*s make triad-ratio judgments,[2] and then from these judgments we generated a multidimensional scale for each individual.[3] We noted that the four color-deficient *S*s (that is, color-deficient in terms of the American Optical Company charts), were distinctly different from the normal *S*s as well as among themselves. In fact, the distortions introduced into the multidimensional space were of varying magnitudes corresponding to the differences in the difficulty which the various *S*s experienced in taking the color-blindneess test. We also noticed some differences between the multidimensional scales for the various normal *S*s. The individual scales seemed to be quite stable as indicated by results from a second series of judgments by some of the individuals. In each case, the second scale was almost identical with the first scale for the individual.

It seems reasonable to suppose that a uniform color-scale would be useful only to the degree that the perceptual color-space has common features across normal *S*s. The data already considered suggest that there are characteristic differences in the way individuals perceive the relations among colors, at least when these perceived distances are measured on a ratio-scale. Perhaps we need several uniform color-scales depending on which type of person is to view the result of the color-measurements. To answer questions such as these, it is necessary to determine the nature and degree of the differences in color-perception for normal individuals. The present research was directed toward this problem and we used the data which already had been collected as part of the earlier study.

Experimental procedures. Ten *S*s with normal color-vision and four with defective color-vision, as determined by the American Optical Company charts, served in the experiment. One of the normal *S*s and one of the color-deficient *S*s were tested twice, with a month intervening between each run. The stimulus-colors were hexagonal tiles, 2 in. across, which had been painted in accordance with specifications developed by the Society Committee on "Uniform Color Scales." Painted with a matte finish, the tiles were of the same lightness and saturation, and equally spaced in hue. The specifications of the stimuli are given in Table I. Thus, in terms of existing scales, the stimulus-configuration would be expected to lie in a plane, and the points representing the stimuli would be expected to lie on a circle and to be approximately uniformly spaced from one another (for normal *Ss*).

---

[2] W. S. Torgerson, *Theory and Methods of Scaling,* 1958, 293.

[3] Gale Young and A. S. Householder, Discussion of a set of points in terms of their mutual distances, *Psychometrika,* 3, 1938, 19-22.

The triads-procedure required that all possible triples of stimuli be presented to *Ss*, which for 10 stimuli totals 120 sets of three. *S* viewed the chips on a neutral gray background and actually moved the chips about on this background in such a way that the physical distances between the chips reflected his perception of the relations among the colors. The analysis of these data provided similarity-values for all possible pairs of 10 color chips. These values were treated directly as interpoint distances. When the matrix of scalar-products was computed, and the characteristic roots and vectors and principal-axis factors were determined for the scalar-product matrices for each *S*, the multidimensional scales for most of the normal *Ss* were, more or less, two-dimensional, and the points representing the stimuli formed a roughly circular configuration in this plane. The scales for the color-deficient *Ss* departed markedly from this pattern and, in general, could be explained in terms of an underestimation of large distances in the red-green direction.

TABLE I

TRI STIMULUS-VALUES, CHROMATICITY-COORDINATES AND THE MUNSELL DESIGNATION OF THE STIMULI

(The first term indicating hue; the second, value (lightness); and the third, chroma (saturation)

Tile	Tri-stimulus-values			Chromaticity		Munsell designation
	X	Y	Z	x	y	H V C
1	3800	3169	4431	0.3333	0.2780	1.3RP/6.0/6.0
2	3505	3165	2680	0.3749	0.3385	10.0R/6.0/4.0
3	3227	3218	2027	0.3809	0.3798	2.5Y/6.0/3.5
4	2804	3174	1860	0.3577	0.4050	3.5GY/6.0/4.0
5	2675	3320	2322	0.3216	0.3992	9.3GY/6.0/4.8
6	2500	3146	3293	0.2797	0.3519	8.8G/6.0/4.4
7	2664	3175	4685	0.2531	0.3017	1.5B/6.0/4.5
8	3003	3166	5593	0.2553	0.2692	1.8PB/6.0/5.6
9	3383	3155	5835	0.2734	0.2550	10.0P/6.0/6.8
10	3642	3189	5488	0.2956	0.2589	5.7P/6.0/7.2

Analysis. The analytical techniques employed represent an application of the procedure developed by Eckart and Young for the approximation of one matrix by another of lower rank,[4] but with one important extension. This extension is embodied in the notion of an 'idealized individual,' which is represented in the analysis as a direction in the factor-space resulting from a factor-analysis of individuals. This permits an analysis of differences between groups of individulas, where the *Ss* within a group share a common perceptual structure and different groups share different structures.

---

[4] Carl Eckart and Gale Young, The approximation of one matrix by another of lower rank, *Psychometrika,* 1, 1936, 211-218.

Results. For the purposes of the present analysis, the interpoint-distance scales for all *S*s (including both scales for the individuals who were tested twice) were arrayed as the X-matrix. The characteristic roots and vectors of the X′X-matrix were next determined. The first three characteristic roots accounted for about 98% of the sums of squares, and there was a distinct break between the third and fourth roots. Three dimensions therefore were retained for further analysis. An inspection of the A-matrix containing the individual coefficients on the principal axes of X showed that the vector corresponding to the largest root was made up of a constant, or very nearly so. These coefficients are related to the approximations of the mean interpoint-distance scale-value for each *S*. A plot of the projections of the individual vectors in the plane of the second and third principal axes may be seen in Fig. 1.

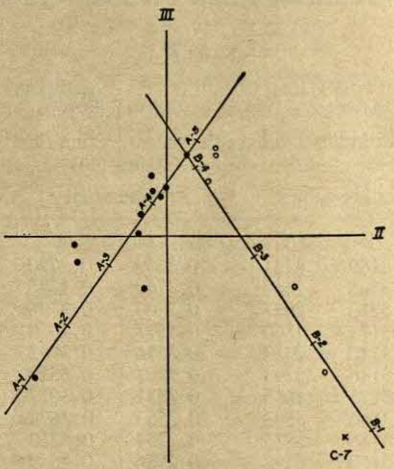


FIG. 1. FACTOR-STRUCTURE OF INDIVIDUAL DISTANCES

The solid points which follow along the A-line all refer to normal *S*s who passed the color-blindness test. The two points rather close to one another, to the left of the line and slightly below axis-2, represent the normal individual who was tested twice. All of the points represented by open circles falling along the B-line represent color-deficient individuals who failed to read all of the charts in the color-blindness test. The two points very close to one another represent the two sets of data for the color-deficient *S* who was tested twice. It is of some interest to note that the person with the severest color deficiency is represented by the point farthest out on the B-line, while the person with the next most serious deficiency is represented by the next point. The point almost on the B-line near the intersection of the two lines represents a person who was unaware previously that he had any color-deficiency. The point right on the intersection of the two lines represents a person classified as normal, but who had some slight difficulty in reading some of the color-blindness charts, and whose multidimensional-scaling results were markedly different from those of the other normal *S*s.

To define more clearly the nature of the differences represented by differences in direction in this space some further analyses were carried out. It is apparent that the columns of the Y-matrix provided approximation to average multidimensional scales which were obtained from individuals who took part in the experiment. We can, however, conceive of idealized individuals who would be represented by other directions in this space of individuals, for example, at the points labelled A1, A2, A3, and so on, in Fig. 1. It is possible to determine the entries corresponding to these points in the matrix of stimulus-pair coefficients by rotating the A-matrix to orient an axis along the direction describing an ideal individual, and then by applying the inverse of this transformation to Y. These transformed entries now can be treated as the scale of interpoint-distances for the idealized person corresponding to the direction. Nine such scales corresponding to the points A1-A5 and B1-B4 were determined and

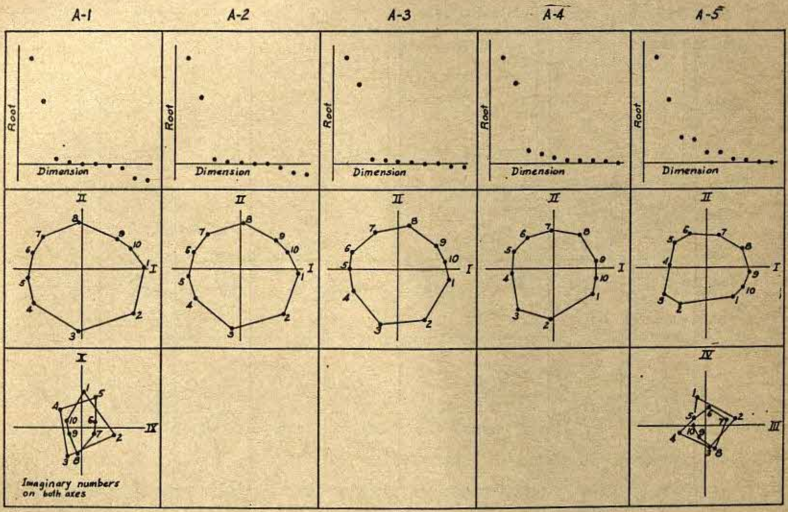


FIG. 2. COLOR-SPACES OF IDEALIZED INDIVIDUALS: LINE A

analyzed multidimensionally. The multidimensional scales which resulted for the A-points may be seen in Fig. 2. It should be noted that the change in the patterns of the characteristic roots is quite systematic as one moves from A1-A5 in the space of individuals, and the direction defined by point A3 provides a multidimensional space that is almost exactly two-dimensional. The vectors corresponding to the *two negative roots* for A1 have highly systematic projections, in that, when adjacent points are connected, a double loop occurs. Other work which we have done shows quite clearly that such a result can occur if *S*s systematically overestimate large distances relative to the small distances. Conversely, the vectors corresponding to the *second largest pair* of positive roots for the space defined by A5 give the same kind of pattern. It can be shown this result will occur if *S*s systematically underestimate large distances relative to the small distances. These results rather clearly suggest that normal *S*s do differ from one another in a highly systematic way in terms of how they report their perception of the relations among colors. Moreover, a uniform color-scale that was appropriate for people in the vicinity of A3, might reasonably be considered inappropriate for persons in the vicinity of A1 or A5. It now becomes a problem to pinpoint the source of the differences between these presumably normal individuals. Do the differences really reflect differences in color-systems or do they reflect a difference in the

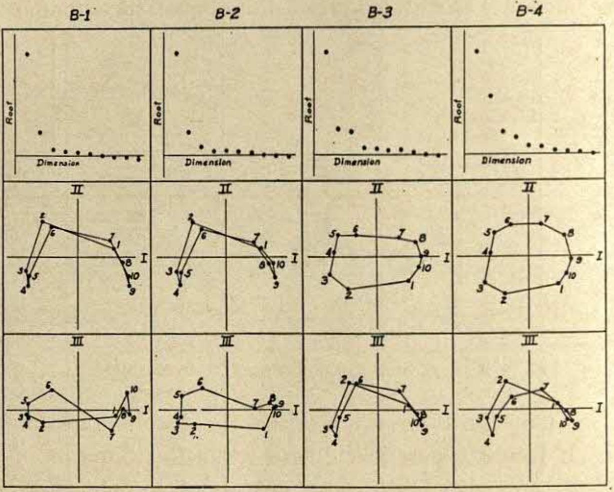


FIG. 3. COLOR-SPACES OF IDEALIZED INDIVIDUALS: LINE B

use of numbers, or in the interpretation of instructions? This is certainly a matter for further research.

Fig. 3 represents the same kind of data for points along the B-line. It may be noted that at point B1 we have a more nearly one-dimensional structure than occurs for directions corresponding to other points along the B-line. Stimuli 2 and 6 are much closer together than for the normal individuals and stimuli 1 and 7 also are much closer together. Stimuli 1 and 2 are a red-purple and a red, respectively, whereas stimuli 6 and 7 are a green and blue-green, respectively. Thus, we observe a marked contraction of distances in the red-green direction of the stimulus-space. As we move in along the B-line we note that the distortion becomes less and less severe. It would appear that differences in this direction of the space of individuals are related to varying degrees of a contraction of one direction of the perceptual space. Since the distortions here are related to particular stimuli, it seems logical to conclude that these differences are principally related to the mechanism of color-perception as opposed to the situation for the normal *S* wherein the distortions were independent of particular stimuli.

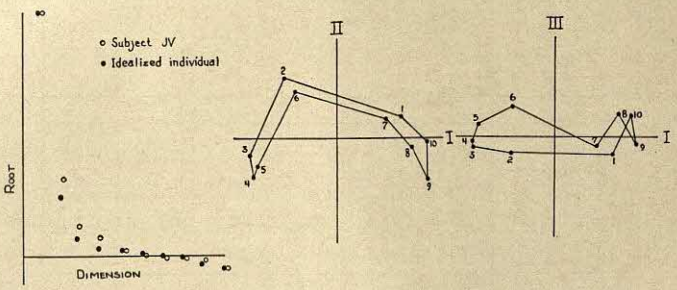


FIG. 4. COLOR-SPACES OF IDEALIZED INDIVIDUAL C-7

Other regions of the individual space also were investigated by the same techniques in an attempt to locate a direction corresponding to an ideal person with a more nearly dichromatic color-system than any belonging to those persons who actually took part in the experiment. Point C7 in Fig. 1 yielded the multidimensional scales which are plotted in Fig. 4. If we compare the relative sizes of the characteristic roots for this scale with those for the most seriously deficient *S* (*JV*), we observe that this result is, in fact, more nearly one-dimensional. It would be quite interesting to find out whether a pure deuteranope would actually fall in this space.

Summary. Torgerson's multidimensional-triad ratio-scaling procedure was utilized to determine similarity-scale-values for all pairs of 10 color-chips for each of 9 normal and 4 color deficient *S*s. These scale-values were arrayed as a matrix with rows for scale-values and columns for the *S*s. The matrix of cross-products between *S*s was factored, yielding a clear three-dimensional simple structure with all the normal *S*s lying in one plane, and all the color-deficient *S*s lying in a second. Factor-scores associated with various directions in this space then were determined and analyzed as multidimensional scales.

THE EFFECTS OF A 'COMPOUND SET' ON LEARNING EFFICIENCY

By PETER K. JANKE, Reed College

A recent experiment by DiVesta and Blake reported evidence for differential effects on learning of three combinations of (a) verbal reinforcement, and (b) knowledge that some principle was the basis for reinforcement.[1] The sets induced by instruction were presumed to permit verbalizations which mediated adaptations to the problem in the form of hypotheses. As might be expected, they found that a set to 'not look for a principle' resulted in making a task involving a principle significantly more difficult to learn than did a set to look for a principle, although similar reinforcement was given in both cases. In short, a set of a positive nature proved more beneficial for learning than did a set of a negative nature.

Following closely the experimental design used by DiVesta and Blake, the present experiment attempts to investigate further the effects of set on learning efficiency. Whereas previous experiments in this field have dealt with relatively simple 'homogeneous' sets, however, the present study deals primarily with a set combining both positive and negative elements—a so-called 'heterogeneous' or 'compound' set.

It was hypothesized that the negative component of the compound set would become less effective with time (as a result of non-reinforcement), allowing the positive component to be reflected in an increase in performance. Thus, since the compound set is assumed to produce a decrement in performance not shared by the positive set, and an increase not shared by the negative set, the prediction was made that it would produce a learning curve which would fall somewhat between the curves produced by the positive and negative sets.

METHOD

Materials. Following DiVesta and Blake, learning materials were adapted from the descriptions of those used by Philbrick and Postman in a verbal-conditioning study.[2] Two-hundred sixteen different stimulus-words varying in length from 2 to

---

* Received for publication November 14, 1960.

[1] F. J. DiVesta and Kathryn Blake, The effects of instructional sets on learning and transfer, this JOURNAL, 72, 1959, 57-67.

[2] E. B. Philbrick and Leo Postman, A further analysis of 'learning without awareness,' this JOURNAL, 68, 1955, 417-424.

10 letters were selected from *The Teacher's Word Book.*[3] One word of each length was then randomly assigned to 24 blocks of 9 words each.

Experimental groups. Four experimental groups (X_1, X_2, X_3, and C) were used. The treatments received by three of these (similar to the three groups used by DiVesta and Blake) differed with respect to (a) whether they were told that a principle was operative, and (b) whether reinforcement was positive or random. The fourth group (Group C) was given the compound set containing positive and negative components, and was given positive reinforcement.

Instructions. Since wording of instructions is crucial to the experiment, instructions will be presented here exactly as they were read to the *S*s in each group. The following instructions were read to Group X_1, given (a) knowledge that some principle was operative, and (b) positive reinforcement.

> I have prepared a group of words and then assigned numbers to each word according to a principle or system. I am interested in seeing how many of these assigned numbers you can determine correctly when you are shown only the word. I will present you with a series of cards. On each card there is typed one word. After you read each word I want you to tell me what number goes with that word. I have used numbers from 1 to 9 inclusive, so you are free to choose from 1 through 9 for each word. There are 216 different words and I will go through the list just once. Give your responses as quickly as you can and be sure to say each word before saying the number. After each response you give, I will say either 'right' or 'wrong.'

Group X_2 was provided a set to 'not look for a principle' but received positive reinforcement. This group received the following instructions.

> You have heard of extra-sensory perception. *S*s are asked to name items which they cannot possibly perceive in any ordinary sense of the word. If they are able to make hits more frequently than they would by chance alone, it is assumed that this is atributable to extra-sensory perception. I am dealing here with the same problem by a method somewhat different from the ones used before. I have prepared a group of words and assigned numbers randomly to each word. I am interested in seeing how many of these assigned numbers you can guess correctly when you are shown only the word. I will present you with a series of cards. On each card there is a word. After you read each word I want you to guess what number goes with that word. I have used numbers from 1 to 9 inclusive, so you can choose from 9 numbers in each case, and with this rather large selection of choices you will naturally not make many hits. There are 216 different words, and I will go through the list just once. If extra-sensory perception exists, you should be able to get more numbers right than you would by chance alone. If it does not exist, you will merely be guessing the number that goes with each word and should not make a score much better than chance. Give your responses as quickly as you can. Be sure to say each word before saying the number. After each response you give, I will say either 'right' or 'wrong'.[4]

The third group, Group X_3, was provided a compound set by (a) being made aware that some principle was operative, but (b) being led to expect a principle different from that actually in operation, by being directed to a specific quality of the stimulus-words (*i.e.* their meaning) different from that on which the reinforcement was actually based (*i.e.* their length). This group was given positive reinforcement and received the following instructions.

---

[3] E. L. Thorndike and Irving Lorge, *The Teacher's Word Book of 30,000 Words*, 1944, *passim.*

[4] These instructions in Groups X_1 and X_2 are modelled closely on those reported by DiVesta and Blake.

> Perhaps you have heard of the term 'concept formation.' This refers to the process of discovering some characteristic that is common to a series of discrete objects, and which sets them apart from other objects. Homer Reed devised an experiment in 1946 whereby he presented his *S*s with cards, on each of which were printed four common words. The cards were grouped so that every card of one group, say, would have the name of an animal among its four words—the other words having nothing in common. To each group Reed gave a nonsense-syllable name. Thus all cards with an animal word among their four were called KUN, cards with a vegetable word, say, would by BEP, and so on. The *S*s were then asked to learn the names of the cards and to discover the meanings of the names assigned to the cards. My experiment is aimed at approaching this same subject from a different angle. I have prepared a group of words and assigned numbers to each word. I am interested in seeing how many of these assigned numbers you can determine correctly when you are shown only the word. I will present you with a series of cards. On each card there is typed one word. After you read each word I want you to tell me what number goes with that word. I have used only numbers from 1 to 9 inclusive, so you are free to choose from 1 through 9 for each word. There are 216 different words and I will go through the list just once. Give your responses as quickly as you can, and be sure to say each word before saying the number. After each response you give, I will say either 'right' or 'wrong.'

A fourth group, Group C, was used as a control, and was made aware that a principle was operative, but was given reinforcement on a random basis. The *S*s in this group received the same instructions as did those in Group X_1.

Principle of reinforcement. For all groups which received positive reinforcement (*i.e.* Groups X_1, X_2, and X_3), the reinforced *E*'s saying 'right' was the number equal to the number of letters in the stimulus-word *minus* one; all other responses were designated as incorrect, and were followed by *E*'s saying 'wrong.' (For example, the only correct response for the stimulus-word 'combat' would be 'five'.) If *S* achieved the level of responding correctly to two consecutive blocks, that is, of achieving at least 18 correct responses in succession, testing was halted and *S* was asked to state the basis on which his responses were being made. If the principle given was incorrect, testing was resumed and the inquiry repeated at the end of each block until *S* verbalized the correct principle, at which point testing was terminated. Both responses and principles stated were recorded by *E* for each *S*.

Scoring procedure. For purposes of statistical analysis, each *S* was scored numerically for his level of performance. One point was given for every correct response; no points for an incorrect response. Thus a perfect score for any block was nine. In cases where *S* verbalized the correct principle before the end of all 216 trials and testing was terminated, nine points were automatically given for every block beyond the point of termination. Total performance scores for all *S*s, therefore, were out of a possible 216 points.

Subjects. The *S*s for this experiments were students at Reed College. No *S* was in any way acquainted with the nature of the experiment before participation. Twenty men were randomly assigned to each of the four groups in such a manner that each group contained five *S*s.

RESULTS

The first question posed by this experiment is: Do differences in instructional sets produce differences in rates of learning? It had been predicted that Group X_1 would make more correct responses than either Group X_2 or X_3; that Group X_3 would perform better than Group X_2; and that Group C, the control with random reinforcement, would show no improvement, remaining near chance.

The performance of each of the four groups is presented graphically in Fig. 1. This figure shows the mean number of correct responses per block for *S*s in Groups X_1, X_2, X_3, and C on each of the 24 blocks (9.0 being perfect score). Group trends are all in the predicted direction. The learning curve for Group X_1 is fairly steep, reaching the highest possible score by the eleventh block of trials. The curve for Group X_3 is roughly parallel to that for Group X_1, and also reaches its peak (7.5) by the eleventh block of trials. Except for one member of Group X_3, all the *S*s in this group reached 9.0 by the eleventh block, as with Group X_1, and the discrepancy between the curves for groups X_1 and X_3 beyond this point is attributable to the one member of Group X_3 who performed at a consistently low level. The curve for Group X_2 is a great deal lower than those for either X_1 or X_3, but there is still some evidence for learning. Group C gives no indication of learning at all; the curve remains flat and not much better than chance.

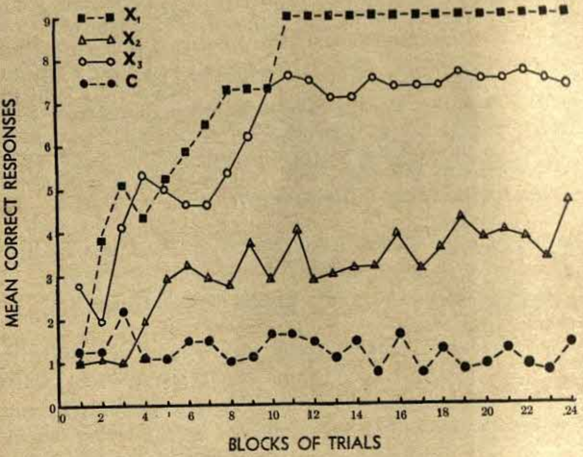


FIG. 1. LEARNING CURVES FOR ALL GROUPS

The significance of these trends was tested by an independent-group analysis of variance over the total trials. Testing the variance among conditions revealed highly significant differences ($F = 10.21$; $p < 0.01$), confirming the results of DiVesta and Blake in this respect. Since this study is particularly concerned with the performance of Group X_3 relative to the other groups, further tests of differences between the means for all 24 blocks were made. To test the prediction that performance for Group X_3 would fall between that for Groups X_1 and X_2, Tukey's method for evaluating differences between independent means following an analysis of variance was used.[5] Results show that both Groups X_1 and X_3 differ from the Control Group beyond the 1% level of significance. In addition, the X_1–X_2 difference is significant beyond the 1% level and the X_2–X_3 difference is significant beyond the 5% level. The other differences failed to reach significance at the 5% level.

DISCUSSION

The performance of *S*s in Groups X_1, X_2, and C provides evidence for the differential effects on learning of three combinations of (a) verbal reinforcement, and (b) knowledge that some principle was the basis for reinforcement, and hence supports the recent findings of DiVesta and Blake. There are, however, certain differences between the earlier data and those reported here. The results for Group X_2 and C agree very closely with the results of the corresponding groups reported by DiVesta and Blake. The performance of Group X_1, however, differs quite markedly from their group given similar conditions. Whereas in the present experiment every member of Group X_1 reached a perfect score of 9.0 by the eleventh block, in the earlier study this group never surpassed a score of 6.0, and only approached this figure toward the twentieth block of testing. Thus, although the relationships among groups remain the same, the results of this experiment indicate more significant differences than those reported earlier.

The primary aim of this experiment, however, was not replication of the earlier study, but, rather, an investigation of the effects of a 'compound set' introduced with the new experimental group, Group X_3. The instructions given Group X_3 are presumed to have aroused a set consisting of two components—one, somewhat implicit, to look for a principle; the other, more explicit, to look for it only on a certain unreinforced basis; namely, word-meaning.

---

[5] As described by T. A. Ryan, Multiple comparisons in psychological research, *Psychol. Bull.*, 56, 1959, 26-47.

Results of the experiment confirmed the prediction that the learning curve for Group X_3 would lie somewhere between the curves for Groups X_1 and for X_2. It is somewhat surprising, however, that the curve for X_3 approached the curve for X_1 as closely as it did. This is reflected statistically by the fact that the difference between the means of X_1 and X_3 failed to reach significance at the 5% level. This result seems to imply that in this case the effect of the negative component was slight, since performance was not significantly inferior to that of the group given a purely positive set. This phenomenon is consistent with the hypothesis concerning the compound set which was outlined above. It was suggested that since the negative component to look for word-meaning was consistently unreinforced, it soon extinguished and left active only the implicit component to look for a principle. Observations made during the experiment also tend to support this hypothesis. The *S*s testing various 'hunches' in the area of word-meaning against the reinforcement were soon observed to extend their hunches beyond the boundaries of this area, and solution of the problem generally followed swiftly. Thus, after a surprisingly brief period of testing, the responses of Group X_3 (with the exception of one *S*) became practically identical with those of Group X_1. The surprising proficiency of members of Group X_3 might be related to intelligence, persistence in problem solving, or a number of related factors. Further investigation of these areas in particular, and of compound set in general, might prove worthwhile in understanding the effects of set on learning.

SUMMARY

This experiment is both a replication and an extension of a recent investigation of the effects of set on learning efficiency performed by DiVesta and Blake. It investigates the hypothesis that various combinations of (a) verbal reinforcement, and (b) knowledge that some principle is the basis for reinforcement, will have differential effects on learning. The effects of a compound set, presumably composed of positive and negative components, were studied.

Twenty men *S*s were asked to respond to 24 blocks of 9 words each with a number from 1 to 9. The *S*s in Group X_1 were instructed that a principle (the number of letters in a word minus one) was the basis for *E*'s saying 'right' or 'wrong' to a response, but were not informed what the principle was. The *S*s in Group X_2 were instructed that *E* was testing the success of extra-sensory perception; those in Group X_3 were informed that a principle was the basis of reinforcement, but were told that the

principle involved the meaning of words. Correct responses of these three groups were reinforced positively. A fourth group, Group C, was informed that a principle was operative, but reinforcement was administered randomly.

Analysis of the data indicates that Group X_1 made more correct responses than Groups X_2 or X_3, that Group X_3 made more correct responses than Group X_2, and that all three experimental groups performed better than Group C, which performed near chance level throughout. Results are interpreted as providing support for DiVesta and Blake's results and as supplying evidence for the operation of a compound set.

ARE 'AUTOPHONIC' JUDGMENTS BASED ON LOUDNESS?

By RICHARD M. WARREN, Shimer College

There have been many psychophysical studies relating loudness-judgments to the physical intensity of sounds heard through headphones or over loudspeakers.[1] In an interesting variation of this procedure by Lane, Catania, and Stevens, *O* produced the vowel 'ä' with his own voice, and at the same time made quantitative judgments of his "vocal level."[2] The authors concluded that these "autophonic" judgments are quite distinct from loudness-judgements involving externally generated sound. Loudness is based upon auditory stimulation, while autophonic judgments, they maintained, are primarily estimates of "muscular effort" employed in producing the sound.

If this conclusion is valid, then it would be anticipated that different autophonic scales would be found for other self-generated sounds involving different degrees of muscular effort and different sets of muscles. In the present study, *O* estimated magnitudes of self-generated sounds which included not only 'ä,' but also the unvoiced fricative consonant 'sh,' and a note produced by blowing on a pitch-pipe. If the same relation between sound-intensity and subjective judgment exists for these different sounds, it would appear that autophonic judgments are primarily auditory rather than muscular in nature.

METHOD

Observers. The 40 *O*s were 18-40 yr. old and had normal hearing. They were unfamiliar with measurement of the physical intensity of sounds, had never served in a previous auditory experiment, and were naïve concerning psychophysical theory.

Apparatus. Sound-intensity was measured on a General Radio 1551A sound-level-intensity-meter using slow response. The musical note of *F* above middle *C* was produced by *O*, blowing on a Master Key Chromatic Pitch Instrument.[3] The note-selector mouthpiece of this pitch-pipe was fixed in the position producing the appropriate note.

Procedure. O was seated in a specially constructed room which attenuated ex-

---

* Received for prior publication April 29, 1962: This research was done at The Laboratory of Psychology, National Institute of Mental Health, Bethesda, Maryland.

[1] For a summary of this literature, see S. S. Stevens, The measurement of loudness, *J. acoust. Soc. Amer.*, 27, 1955, 815-829.

[2] H. L. Lane, A. C. Catania, and S. S. Stevens, Voice level: autophonic scale, perceived loudness, and effects of sidetone, *J. acoust. Soc. Amer.*, 33, 1961, 160-167.

[3] Manufactured by Wm. Kratt Company, Union, New Jersey.

terior sounds >30 db. The ambient level inside this room was <25 db. *O* was requested to produce a sound at a "moderate level" and to hold it steady at this level for several seconds. He was told that *E* would signal him when he had kept it steady long enough. He then was to take another breath, and immediately to produce the sound at half the previous level, again holding it steady until signaled by *E*. To qualify for *E*'s signal, *O* was required to hold intensity at a level deviating less than 1 db. from a central value for 3 sec. Sound-intensity was measured 6 ft. in front of *O*'s head, with the dial of the meter out of his sight.

If both the first, 'moderate' level and the second, 'half-level' produced by *O* met the steadiness-criterion described above, they were considered as a half-level judgment. Three such judgments were made for each type of self-generated sound. Twenty *O*s made three judgments of half-level, first with 'ä,' and then with 'sh.' The other 20 *O*s reversed this order, first producing the sound 'sh,' then 'ä.' Ten *O*s judging the phonemes in the order 'ä' then 'sh,' and 10 *O*s judging the phonemes in the order 'sh' then 'ä,' made three final half-level judgments of the musical note they produced by blowing through the pitch-pipe. Thus, for both phonemes 'ä' and 'sh,' 40 *O*s each made three autophonic judgments (a total of 120 judgments for each sound). For the pitch-pipe note, 20 *O*s each made three autophonic judgments (a total of 60 judgments).

TABLE I

INTENSITY-RATIOS (EXPRESSED AS DB-ATTENUATION) CORRESPONDING TO HALF-LEVEL JUDGMENTS FOR SELF-GENERATED SOUNDS

	Median	Mean	SE
Vowel 'ä'	7.0	7.35	0.29
Unvoiced consonant 'sh'	7.0	7.43	0.25
Pitch-pipe note	8.0	8.05	0.51

RESULTS AND DISCUSSION

The ratio of sound-intensities (expressed in db.) corresponding to judgments of half-level are given in Table I for 'ä,' 'sh,' and the note of a pitch-pipe. The differences between means for the different sounds are not significant in any case. Because of the skewness of distribution of responses, the median appears a better measure of central tendency than the mean. The mode for the distribution of judgments was 6 db. for all three sounds. These results indicate that autophonic judgments are based upon sound-intensity rather than amount of muscular effort.

Different sets of muscles, different efforts, different physical principles underlie production of the three self-generated sounds employed in the present study. Let us consider briefly the mechanism for production of these sounds. Speech-production involves complex coördination of impulses from the 5th, 7th, 9th, 10th, 11th, and 12th cranial nerves.[4] In the case of the phoneme 'ä' (the vowel found in the word *tar*), several muscles of the chest and abdomen produce pressure expelling air past the vocal cords in the larynx. The vocal cords, when moved to the appropriate position through muscular activity, periodically open and close, giving rise to the changes in pressure which constitute sound. When the intensity of the vowel is increased, not only do thoracic and abdominal muscles exert greater pressure in expelling air, but the intrinsic muscles of the larynx change the type of movement of the vocal cords—closing them more firmly and opening them for a shorter proportion of the time.[5]

---

[4] R-M. S. Heffner, *General Phonetics*, 1949, 41.

This control of intensity by laryngeal muscles is absent for the phoneme 'sh' (the consonant found in the word *she*). In producing this sound, the vocal cords are kept wide open and do not contribute to sound-production. The tongue is moved close to the hard palate, and the turbulent passage of air through this small opening gives rise to this unvoiced fricative consonant. Movements of the tongue reducing or increasing the size of this orifice are used to control intensity of 'sh' by *O*s, along with changes in pressure of the expelled air. This control of intensity by the tongue involves muscles, to be sure, but control is in terms of shape and position of the tongue rather than by 'muscular effort.'

While the phonemes used in this study must have been generated by the *O*s many millions of times, none had experienced producing sound with the pitch-pipe, and therefore they were allowed to 'try it out' for about 15 sec. before making psychophysical judgments. The intensity of sound produced by the pitch-pipe was controlled by *O* solely by pressure of the expelled air. The sound was produced by air passing over a flexible metal strip fixed at one end. Vibration of this strip increased and decreased in a periodic fashion the size of the orifice through which air escaped—giving rise to a musical note. The relation between air-pressure at the mouthpiece and the intensity of sound produced was a complex function characteristic of the particular instrument employed. Moderately hard blowing produced the maximal sound-level; blowing harder did not affect sound-level.

The fact that no statistically significant difference was found between the ratio of sound-energies corresponding to half-intensity-level judgments with the three different self-generated sounds, each involving different sets of muscles and different manners of setting air in vibration, is consistent with the concept that *O* bases his judgments on the loudness of sound produced. This equivalence of attenuation would not be anticipated from the theory that autophonic judgments are not based primarily upon hearing, but rather upon estimates of muscular effort. Indeed, it would seem scarcely possible that these different muscular efforts should produce equivalent db.-attenuations without the mediation of hearing.

Lane, Catania, and Stevens gave two major reasons for concluding that hearing plays a minor role in quantitative estimates of *O*'s own voice-level. One reason was that essentially the same autophonic scale was obtained whether or not masking noise or sidetones of different intensities were heard through headphones during *O*'s vocalization. Since, however, *O* is quite familiar with production of speech-sounds, it may be that experimental interference with auditory monitoring of vocal level brings into play compensatory mechanisms for control. The close relation between hearing and speech-production is well known. Indeed, congenital difficulties in hearing frequently are indicated first by a child's inability to produce speech having proper qualitative and intensitive relations of phonemes. Complete congenital deafness is associated with mutism. The ability of a normal *O* to speak and produce phonemes of proper intensity and quality even if temporarily deafened does not imply that hearing is not of primary importance in the genesis and maintenance of normal vocal control.

---

[5] Harvey Fletcher, *Speech and Hearing in Communication,* 1953, 22-23.

The other reason given for minimizing the role of hearing in autophonic judgments was that differences were found between the psychophysical functions for autophonic 'ä' and for 'ä' produced by someone else. According to the functions reported, 8.6 db. corresponds to half-loudness for externally produced 'ä' compared to 5.5 db. for half-vocal-level for *O*'s own voice. The difference between these functions was cited as direct evidence that autophonic judgments are not based upon hearing. These results do not, however, necessarily imply a non-auditory basis for autophonic judgments. Indeed, the autophonic scale is in quantitative agreement with a theory concerning loudness of externally produced sound—the physical-correlate theory.[6] According to this theory, the manner in which sound varies with distance from the source serves as the basis for loudness-judgments. From physics, an attenuation of 6 db. corresponds to doubling distance, but changing distance also involves changes in quality (*e.g.* degree of reverberation). Warren, Sersen, and Pores found that half-loudness of speech coming from a fixed external source corresponded to an attenuation of 7.9 db. (in good agreement with Lane, Catania and Stevens' value of 8.6 db. for externally generated 'ä') and presented evidence indicating that the deviation of this value from the theoretical 6 db. was due to conflict between changing intensitive and fixed qualitative correlates of distance.[7]

Self-generated sounds represent unique, highly interesting sounds from the point of view of the physical-correlate theory. With autophonic judgments, the influence of qualitative cues to distance are eliminated for the speaker (although not for his audience). In normal speech, the intensity of sound generated depends upon the distance from the speaker to the person being addressed. We shout to a distant person, speak in a low voice to someone nearby, and use the appropriate vocal levels for intermediate distances. If the distance from speaker to listener serves as the physical correlate for autophonic judgments of intensity, it would be expected that judgments of half-vocal-level would approximate the attenuation corresponding to projecting one's voice for half the distance—a decrease of 6.0 db.

---

[6] R. M. Warren, A basis for judgments of sensory intensity, this JOURNAL, 71, 1958, 675-687.

[7] R. M. Warren, E. A. Sersen, and E. B. Pores, A basis for loudness-judgments, this JOURNAL, 71, 1958, 700-709.

According to the autophonic scale, the half-vocal-level of 'ä' corresponds to an attenuation of 5.5 db. Clark and co-workers found that successive halving of vocal level of 'ä' averaged 6.0 db.[8] In the present study, an attenuation of 7.0 db. corresponded to half-vocal-level for the same phoneme. The average of the three studies is 6.2 db.—approximating the value of 6.0 db. derived from the physical-correlate theory.

Hearing is a distance-sense, and we are continually aware of spatial relationships while listening to sounds or when producing sounds ourselves for an audience. It appears that this intimate correspondence between hearing and spatial localization is responsible for quantitative judgments of the magnitude of sounds, whether self-generated or externally produced.

SUMMARY

In a recent study, Lane, Catania, and Stevens required *O* to produce the vowel 'ä' at various intensities, and to make quantitative judgments of relative "vocal level." They concluded that these 'autophonic' judgments of self-generated sounds were based primarily upon estimates of muscular effort rather than upon loudness.

In the present study, autophonic judgments were obtained for 'ä,' for the unvoiced consonant 'sh,' and for the sound produced by blowing on a pitch-pipe. The attenuation (in db.) corresponding to half-level was equivalent for these three sounds, despite their different relations to muscular effort. It is suggested that autophonic judgments are loudness-judgments. A theory proposed previously for loudness appears to apply as well to autophonic judgments.

---

[8] This datum (by K. C. Clark, H. W. Rudmose, J. C. Eisenstein, F. D. Carlson, and R. A. Walker, The effects of high altitude on speech, *J. acoust. Soc. Amer.*, 20, 1948, 776-786) is cited by Lane, Catania, and Stevens, *op. cit.*, 162.

KINESTHETIC AFTER-EFFECTS AND LENGTH OF INSPECTION-PERIOD

By Paul Bakan, Louis B. Myers, and James Schoonard, Michigan State University

A number of studies on the temporal aspects of figural after-effects have been reported.[1] Most of them have been concerned with visual after-effects, few with kinesthetic. Such evidence as is available, however, indicates that kinesthetic after-effects increase with repeated inspection-periods but the increases have not been studied as a function of length of a single inspection-period.[2]

The aim of this investigation is twofold: (1) to study the growth of the kinesthetic after-effect with increasing duration of the inspection-period and (2) to determine the decay of the after-effect as a function of the length of the inspection-period.

Method. The *Ss*, 80 undergraduate volunteers, were divided into four groups of 20 each: a control group and three experimental groups. The lengths of the kinesthetic inspection-periods of the four groups differed. The three experimental groups were allowed 10, 30, and 60 sec. respectively, for inspection; the control group served without an inspection-period.

The apparatus consisted of three wooden blocks: (1) a standard with parallel sides, 1.5 in. wide, 36 in. long, and 2 in. thick; (2) a wedge-shaped block, 28 in. long, and 2 in. thick, which varied in width from 1 to 2⅛ in. at the rate of 1/24 in. per inch; and (3) an inspection-stimulus, with parallel sides, 0.5 in. wide, 24 in. long and 2 in. thick. The three blocks were mounted on tables 34 in. high and parallel to each other. The standard- and inspection-blocks were at *S*'s left and the variable at *S*'s right. The variable block was fitted with an H-shaped rider within which *S* placed his thumb and forefinger while making equality-matches.

S held the standard block between the thumb and forefinger of his left hand, while with his right hand he found a point on the variable block that he deemed equal in width to the standard. For the experimental groups, measures of the point

---

* Received for publication January 3, 1961. This research was supported by Grant G5822 from the National Science Foundation. The assistance of Mr. Richard W. Thompson in the analysis of the data is gratefully acknowledged.

[1] J. F. Bales and G. L. Follansbee, The after-effect of the perception of curved lines, *J. exp. Psychol.*, 18, 1935, 499-503; J. J. Gibson and Minnie Radner, Adaptation, after-effect and contrast in the perception of tilted lines: I. Quantitative studies, *ibid.*, 20, 1937, 453-467; E. R. Hammer, Temporal factors in figural after effects, this Journal, 62, 1949, 337-354.

[2] Wolfgang Köhler and Dorothy Dinnerstein, Figural after-effects in kinaesthesis, in *Miscell. Psychologica Albert Michotte*, 1947, 196-220; Michael Wertheimer and C. M. Leventhal, "Permanent" satiation phenomena with kinesthetic figural after-effects, *J. exp. Psychol.*, 55, 1958, 255-257.

of subjective equality (*PSE*) were taken before and after exposure to the inspection-stimulus. Between these measures, *S* moved the thumb and forefinger of his left hand back and forth along the inspection-stimulus for 10, 30, or 60 sec. depending on the experimental group to which he belonged.

Before *S* was introduced to the apparatus, he was shown a schematic drawing of the standard and variable blocks and was told he was to find a width on the variable which by touch appeared equal to that of the standard. He was then blindfolded, led between the standard and the variable blocks and allowed to explore the entire length of the variable. Two practice trials were allowed: one starting from a point 8/24 in. narrower than standard and one from a point 8/24 in. wider than standard. He was encouraged to move his hand back and forth to 'zero in' on the *PSE*.

An outline of the procedure following the practice trials follows. (1) Pre-inspection judgments: four judgments; two ascending (A) and two descending (D), in ADDA order; mean of four judgments is the *PSE*. (2) Inspection-period: either 10, 30, or 60 sec. for the experimental groups. The *S*s in the control group were not exposed to the inspection-block but spent either 10, 30, or 60 sec. with hands flat on the table and fingers together. (3) Post-inspection judgments: determination of the *PSE* as in (1). (4) Rest-period: *S* keeps hands down on the table, with fingers together, until the start of first judgments of retention. (5) Judgments of retention: Determination of the *PSE*, as noted above, at 5, 10, and 15 min. after conclusion of inspection-period, without further inspection.

This procedure produced four measures of the *PSE*, one immediately following inspection, and the others at successive 5-min. intervals. The differences between the pre-inspection *PSE* and each of the post-inspection *PSE*s were used as measures of the immediate and the retained after-effects.

Results. The curve showing the growth of the after-effect as a function of length of the inspection-period is shown in Fig. 1. The solid line represents the results for the measure of the *PSE* immediately following inspection. The dotted line is based on the average of all four post-inspection *PSE*s.

There is generally an increase in the size of the after-effect with increasing length of inspection-period within the 0- to 60-sec. range.[3] These results are in line with those of Gibson and Hammer for visual after-effects.[4]

An analysis of variance was made and the results show that the values of *F* are significant at the 1% level for (a) length of inspection-period and (b) the interaction between length of inspection-period and successive post-inspection trials.[5]

---

[3] Data from an unpublished study indicate that the size of the after-effect after a continuous 5-min. inspection is not significantly greater than it is after 60 sec. of inspection in this study.

[4] Gibson and Radner, *op. cit.*, 453-467; Hammer, *op. cit.*, 337-354.

[5] The analysis used is the "repeated measurements analysis" suggested in A. L. Edwards, *Experimental Design in Psychological Research*, 1950, 288-296.

Analyses by *t*-test, using error-terms from the analysis of variance (after establishing that there was homogeneity of variance) indicate that the difference between post-inspections 0 and 60 sec. and 10 and 60 sec. are, for the first trial, significant at the 1% level. For the subsequent trials (5, 10, and 15 min. after inspection) the difference between 0 and 60 sec. of inspection is the only one significant at the 1% level.

The nature of the interaction is evident from the set of curves in Fig. 2. For both the 0- and 10-sec. inspections there is an increase in the *PSE* from the first measure to later measures at 5-min. intervals. Under both of these, the difference between the first *PSE* and each of the other three measures of *PSE* is significant at the 5% level.

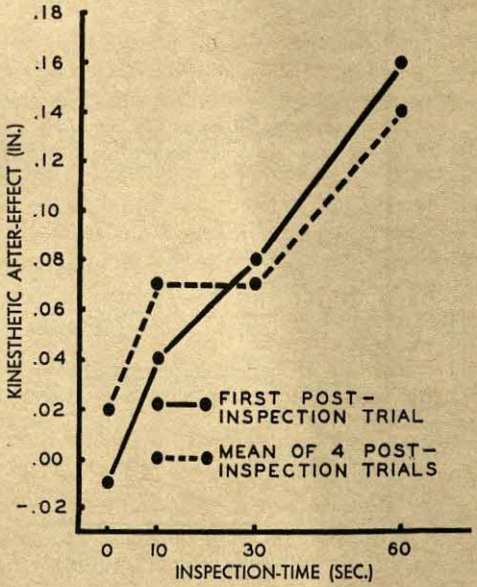


FIG. 1. SIZE OF KINESTHETIC AFTER-EFFECT AS A FUNCTION OF LENGTH OF THE INSPECTION-PERIOD

For the 30- and 60-sec. inspections there is a tendency for the after-effect to decay with time. The differences between the first and fourth trials are just short of significance at the 5% level. The difference between the first and third period for the 60-sec. inspection is significant at the 5% level. Despite this decay, there remains a residual after-effect 15-min. after inspection.

Discussion. This experiment indicates a relationship between size of after-effect and length of the inspection-period. In this respect, the data parallel those obtained from studies on visual after-effects. The results do not, however, parallel the visual after-effects with respect to decay. Hammer found complete decay in 90 sec., following a 60-sec. inspection-period.[6] In the present experiment, there is a significant residual after-effect 15 min. after inspection for the 60-sec. inspection.

The significant interaction between length of inspection-period and post-inspection trials appears somewhat paradoxical. The 10-sec. inspection, instead of showing a decay, shows an increase. A possible explanation would be that a short inspection-period shows a delayed effect, but this is contradicted when the similarity between the 10-sec. and the control curve is considered. The control group with no inspection also shows an increase in *PSE* over time.

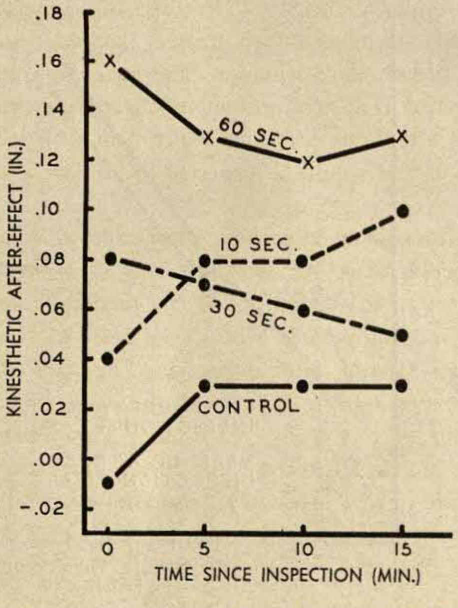


FIG. 2. SIZE OF KINESTHETIC AFTER-EFFECT AT DIFFERENT INTERVALS AFTER THE INSPECTION-PERIOD

These paradoxical increments over time could possibly be due to what Köhler and Dinnerstein have called self-satiation.[7] They found that holding a stimulus-block can by itself produce a change in its apparent width without the interpolation of an inspection-stimulus. For widths of from 0.5 to 3 in. they report that an increase in apparent width is the direction taken by self-satiation. This is the direction taken by the control and 10-sec. groups in the present experiment. It is suggested that self-satiation occurring during the measurement of the *PSE* might help to account for the paradoxical increase in *PSE* on repeated measures for the same individuals.

---

[6] Hammer, *op. cit.,* 337-354.

[7] Köhler and Dinnerstein, *op. cit.,* 196-220, Self-satiation in the visual modality

If self-satiation is operating then it is necessary to consider the possibility that the results of the 30- and 60-sec. inspections reflect interaction between satiation from exposure to the inspection-stimulus and self-satiation from the standard stimulus. The interaction of self-satiation on each post-inspection trial may prevent more rapid decay over time. If this were true, then a replication of this experiment with independent groups on each retention-trial would be expected to produce a steeper decay.

Summary. Measures of kinesthetic after-effect were taken with length of inspection-period of 0, 10, 30, and 60 sec. The size of the after-effect immediately after an inspection-period is a monotonic increasing function of length of inspection-period. Measures of residual after-effect were also taken 5, 10, and 15 min. after inspection. The after-effects after 30 and 60 sec. of inspection tend to decay with the passage of time, but do not, within the limits of this study, reach zero. The apparent width of the standard block increases after 0- and 10-sec. inspection. This is in the direction opposite to that of decay of the after-effect and results in a significant interaction between length of inspection-period and post-inspection trials. An attempt was made to explain these findings in terms of self-satiation.

---

has been investigated by C. P. Duncan (Effect of self-satiation on perceived size of a visual figure. *J. exp. Psychol.*, 60, 1960, 130-136) and H. Ikeda and T. Obonai (The quantitative analysis of figural after-effects: III. On "self-satiation," *Jap. J. Psychol.*, 24, 1953, 179-192).

UNDERESTIMATION IN SIZE-CONSTANCY JUDGMENTS

By V. R. CARLSON, National Institute of Mental Health

The thesis which suggests the present research is that size-constancy represents the fundamental, natural relationship of perceived size to object-size. Deviations from this relationship are produced in experiments on size-constancy through the operation of extra-perceptual response-factors. Overestimation of object-size as a function of distance was previously derived on this basis.[1] The present study is an application of the same hypothesis to the underestimation of the size of objects.

That appropriate instructions can produce underestimations of size is well-established.[2] An alternative method of manipulating the psychological state of *O* appeared to be possible in the administration of d-lysergic acid diethylamide (*LSD*), a drug which acts centrally. This drug may produce a wealth of subjective effects without necessarily impairing actual perceptual-motor performance, and the correlation between subjective manifestations and more objective measures of effect appears to be low.[3] These discrepancies suggest that a primary psychological effect of *LSD* is a shift in *O*'s orientation of report toward the subjective,[4] and, if that is so, it is predicted that he will underestimate the size of a more distant object relative to a nearer one.

---

* Received for publication January 6, 1961. The author is greatly indebted to Mr. Eugene Tassone and Mr. William Jones for assistance in the collection and analysis of the data.

[1] V. R. Carlson, Overestimation in size-constancy judgments, this JOURNAL, 73, 1960, 199-213; Effects of sleep-deprivation and chlorpromazine on size-constancy judgments, this JOURNAL, 74, 1961, 552-560.

[2] Egon Brunswik, *Perception and the Representative Design of Psychological Experiments,* 1956, 96-99; Carlson, *op. cit.,* this JOURNAL, 1960, 199-213; A. S. Gilinsky, The effect of attitude upon the perception of size, this JOURNAL, 68, 1955, 173-192; B. E. Holaday, Die Grössenkonstanz der Sehdinge bei Variation der inneren und äusseren Wahrnehmungsbedingungen, *Arch. ges. Psychol.,* 88, 1933, 449-452; Nöel Jenkin and Ray Hyman, Attitude and distance-estimation as variables in size-matching, this JOURNAL, 72, 1959, 68-76; J. L. Singer, Personal and environmental determinants of perception in a size constancy experiment, *J. exp. Psychol.,* 43, 1952, 420-427.

[3] Conan Kornetsky, Relation of physiological and psychological effects of lysergic acid diethylamide, *A. M. A. Arch. Neurol. Psychiat.,* 77, 1957, 657-658; Conan Kornetsky, Ogretta Humphries, and Edward Evarts, Comparison of psychological effects of certain centrally acting drugs in man, *ibid.,* 77, 1957, 318-324; Abraham Wikler, *The Relation of Psychiatry to Pharmacology,* 1957, 69-78.

[4] Sidney Cohen, Lionel Fichman, and Betty Eisner, Subjective reports of lysergic acid experiences in a context of psychological test performance, *Amer. J. Psychiat.,* 115, 1958, 31.

On the basis of this argument, preliminary data were collected for a number of *O*s after administration of *LSD* on one occasion and administration of placebo on another. An over-all comparison between the condition produced by *LSD* and the placebo did not show a statistically significant difference, but the scores of the *O*s for whom *LSD* was given first did show a difference in the predicted direction. The purpose of the present experiment, with a new sample of *O*s, was to see whether the earlier finding on the effect of order of conditions could be replicated.

Procedure: Observers. Twenty men (18–25 yr. old) served as the *O*s. They were chosen from a population of normal volunteers for medical research. The selection was a routine screening for absence of medical or psychiatric contraindications to participation in drug-studies. They were assigned at random to two groups of 10 each, treated identically except that one group received *LSD* on the first occasion and a placebo on the second; the other group received a placebo first and *LSD* second. None had prior experience either with *LSD* or with size-constancy. If *O* normally wore glasses, he wore them during the tests.

Drug-administration. Seventy-five micrograms of *LSD* were administered orally. The placebo was administered in the same way in an identical capsule. *O* was told beforehand that he would receive a drug on both occasions but not what it would be. *E* did not know whether *LSD* or placebo had been administered in any given instance. Testing was begun 1 hr. after drug-administration, and 3–7 days intervened between successive conditions.

Experimental tasks. The apparatus and test have been described previously in greater detail.[5] Essentially, *O* was required to adjust a variable triangle at a distance of 10 ft. to match the size of a standard triangle at 40 ft. The lateral separation between the two was 20°. Eleven values of the standard triangle (76–159 mm. in altitude) were presented randomly twice each at each session. After completing the size-constancy judgments, *O* was tested with a Bausch and Lomb Ortho-Rater. The standard test-battery was used, comprising determinations of binocular and monocular acuity, lateral and vertical phoria, and stereopsis.

Instructions. With both *LSD* and the placebo, the following instructions were given:

> Adjust the variable so that it looks equal to the standard in apparent visual size. It may also be equal in actual physical size at that point, or it may not—we are not concerned about that. Try to adjust the variable so that it appears equal to you visually, whether you think it is equal in actual size or not.[6]

This instruction is somewhat different from apparent-size instructions used by other investigators in that it does not imply that apparent size must be something different from objective size. It has generally, though not always, resulted in group-averages which do not differ significantly from objective matches.

---

[5] Carlson, *op. cit.* this JOURNAL, 73, 1960, 201-203.

[6] Due to an error, this instruction appeared with a slight difference in wording in a previous report (Carlson, *op. cit.*, this JOURNAL, 73, 1960, 202). The wording given here was used in both cases.

Results. The size-constancy score of each *O* was computed as the average ratio of his setting of the variable triangle to the presented size of the standard triangle for the three main sizes of the standard (76, 117, and 159 mm.). The means of these scores are shown in Table I. The *O*s who received placebo on the first occasion did not, on the average, deviate significantly from an objectively equal match in either the placebo-condition or the LSD-condition, but those *O*s who receved *LSD* on the first occasion significantly underestimated the standard in both conditions.

The sizes in the middle of the range were used to compute a measure of within-*O* precision. These relatively small variations in the size of the standard triangle were significantly reflected in the *O*s' responses but no more so in one group or condition than another.

TABLE I

AVERAGE RATIO OF *O*S' SETTINGS

		Condition	
Group	Measure	*LSD*	placebo
LSD first	Mean	0.70*	0.69*
$(N=10)$	*SD*	0.20	0.18
Placebo first	Mean	1.09	1.00
$(N=10)$	*SD*	0.24	0.18
	p†	<0.01	<0.01

* Significantly less than 1.00, *t*-test, $df=9$, $p<0.01$.
† Probability associated with *t*-test between uncorrelated means, $df=18$.

LSD produced a consistent, slight decrement in Ortho-Rater performance for both groups of *O*s, but there was no significant differential effect dependent upon order of receiving the drug. Correlations between the Ortho-Rater and size-constancy scores also were computed, with no relationship emerging at a statistically reliable level.

Discussion. The major result of this investigation is that *LSD* produced relative size-underestimation only when the drug-condition represented *O*'s first experience with the experimental task. This effect cannot have been a necessary and direct one on any basic sensory process or capacity, for the effect was present or absent independently of the drug itself at a second testing session. The conclusion seems inescapable that different response-sets or approaches to the task were established in the first session, depending upon whether *LSD* or the placebo had been administered; whichever set was established first was exhibited on the second occasion.

This finding actually is not very different from what happens in a simple repetition of the task with no difference in conditions. A reasonably high degree of reliability has been established for judgments of size-constancy obtained with the present procedures. Some *O*s, for whatever reasons, underestimate size at a first session, while others overestimate it. Those who underestimate initially tend to underesitmate subsequently, while those who overestimate initially tend to overestimate the second time. Such consistency is the rule for each of several different instructions which produce clearly differing averages (perfect constancy, overconstancy, or underconstancy) for a group of *O*s as a whole.

According to the present point of view, the general response-bias responsible for systematic deviations from size-constancy under non-impoverished experimental conditions is the perespective attitude, defined as the implicit assumption on the part of *O* that objects become smaller in subjective appearance as a function of increasing object-distance.[7] To be equal in subjective appearance, a nearer object must be smaller in physical size than a more distant one. How much smaller is purely a matter of how stronly *O* is motivated to judge in terms of 'apparent,' as opposed to 'objective,' size. Administration of *LSD* seems to influence *O* in the subjective direction, as predicted; but only if he has not already performed the task under more nearly normal circumstances.

Summary. Administration of *LSD* was used as a vehicle for altering judgments in a size-constancy task. One group of *O*s received a placebo on the first experiment and *LSD* on the second. Another group received *LSD* first and a placebo second. The group receiving the placebo first, produced an average size-match which was not significantly different from size-constancy on either occasion. The group receiving *LSD* first, exhibited underconstancy on both occasions. Measurements of visual acuity, phoria, and stereopsis also were obtained at each session. *LSD* affected these visual functions but not differentially for the groups receiving the placebo and *LSD* first. It is believed that underconstancy is due to the operation of a response-bias based upon the perspective attitude.

---

[7] J. J. Gibson, The visual field and the visual world: A reply to Professor Boring, *Psychol. Rev.*, 59, 1952, 149-151; Carlson, *op. cit.*, 73, 1960, 199-200.

INCREASING PROBABILITY OF INITIAL RECALL WITH ADDITIONAL PRACTICE

By ARCHIE B. CARRAN, Beloit College

In a series of studies, Rock has presented evidence from which it might be concluded that the initial formation of paired associates occurs at full strength in one trial.[1] Rock's two basic procedures were to retain all paired items on subsequent trials for one group but to substitute new paired items for missed ones in another, the substitution group. In numerous variations on this procedure the two groups did not differ significantly in their learning rates. If the above conclusion is correct it has profound implications in favor of theories which maintain that learning occurs in one trial and not gradually.

The writer suspected that Rock's results are only a special case which might be included within a theory of gradual learning. In a previous study, a mathematical theory was developed which related excitatory and inhibitory factors, and the number of repetitions of practice, to the probability of initial correct recall.[2] The theory suggested operations that were used in a quite indirect demonstration of gradual initial acquisition of paired associates. The present study further investigates the relation of practice, which is insufficient for recall of a paired associate item, to the probability of correct recall after additional practice. A more direct approach which is analogous to Rock's design, is used but it incorporates specific features suggested by the theory. The theory suggested, and it is readily appreciated, that the likelihood of demonstrating an increase, rather than a possible *decrease* or constancy, in the probability of an initial correct recall with additional practice is maximized by using the following general operations:

(1) Exposure of a pair of items during practice as briefly as is feasible. It is assumed that the strength of an associative connection grows to an asymptote as a function of the number of implicit rote readings by *S*. Obviously the possible *additional* increment to the probability of correct recall should tend to be greater when briefer exposures are used. The same reasoning applies in operations (2) through

---

* Received for publication January 16, 1961.

[1] Irvin Rock, The role of repetition in associative learning, this JOURNAL, 70, 1957, 186-193; Repetition and learning, *Scient. Amer.*, 199, 1958, 68-72; Irvin Rock, and Walter Heimer, Further evidence of one-trial associative learning, this JOURNAL, 72, 1959, 1-16.

[2] A. B. Carran, A test of gradual versus one trial theories of learning of paired associates. Doctoral Dissertation, University of Cincinnati, 1960.

(4) by means of which the attempt is made to decrease the size of the fraction of all possible connective strength attainable which is acquired in the prior practice. Then a relatively greater additional strengthening will occur with subsequent practice, overcome any more slowly acquired inhibitory effects, and thus cause an increase in the probability of initial correct recall.

(2) A pair of items is removed when it is practiced and missed twice.

(3) *S* is busied with extraneous activities of low interference, when not practicing or attempting recall, to suppress any attempts at implicit rehearsal of paired items.

(4) Difficult items are used. Possibly easy items are easy because *S* has formed correct associative connections of appreciable strength before serving.

(5–6) Items chosen are as dissimilar as feasible and practice of a particular pair of items is separated from practice of others. Both of these operations may decrease the amount of inhibition and interference which accrues with additional practice and may thereby suppress the effect of increased strength of correct associative tendencies due to additional practice.

The following study was definitely an attempt to provide evidence of an increase in the probability of initial recall with additional practice and hence support a theory of gradual initial learning as distinct from full strength acquisition in only one trial. A two-tailed test is appropriate since the perfect dissimilarity of all items and their equality with respect to difficulty may not have been achieved.

METHOD

Subjects. The *S*s were 40 volunteers from introductory classes in psychology.

Materials. Special equipment included a table with shoulder-high two-way screen, an electric metronome, a card holder, and 1 × 2.5-in. cards. The pairs of items used, shown in Table I, are modifications of items published by Woodworth and Schlosberg.[3]

Procedure. The *S*s were paired. A non-substitutive procedure was followed with the first *S* of each pair (S_1). Six paired items were randomly chosen for his first trial. Pairs missed were repeated on a second and final trial. The six paired items not used on the first trial were each assigned at random to one of the paired items used on the first trial. If *S* recalled a particular response-item even partially correctly on his first trial, its mate was substituted in its place for the second trial. For the second *S* (S_2) of each pair, the substitutive procedure was followed. He was given a second trial which was identical in all respects to the second trial for S_1.
The remaining six paired items were used for the paired items on his first trial. Hence the *S* (S_1), who was given the non-substitutive procedure, practiced and was tested again on paired items missed completely in the first trial and his pair mate, (S_2), who was given the substitutive procedure, received only one practice and testing on the same paired items.

S was seated across the table from *E,* who read the instructions to him. Then *S* was given at least two practice trials with digit-items and was explicitly assigned the responsibility for not rehearsing. After this practice, *E* began the first of two trials by manually presenting the cards, masked by a blank card, and saying "ready." The first stimulus-card (the pair of items) was exposed to *S* for 1 sec. Then it was removed and *S* counted backwards aloud for 5 sec. from a three-digit number, exposed by the stimulus-card's removal. With the removal of the number-card, the next stimulus-card (the next pair of items) was exposed. This procedure was continued until the last pair was removed and *E* said "write please." *S* wrote the digits 1 through 9, saying each aloud, during the 15-sec. interval before testing for recall. Then *E* said "ready" and presented the first stimulus-item alone for 7 sec. regardless of *S*s behavior. The next stimulus-item was exposed for 7 sec. by the removal of the first item, and so on. Immediately after testing on the sixth item, *E* again said "write please" and *S* repeated the 15-sec. task before beginning the final trial. The order of paired items and of stimulus-items in testing for recall was varied by shuffling the cards for 4 sec. during the 15-sec. periods. This could be done in advance for the first as well as the second trials with the substitutive procedure.

---

[3] R. S. Woodworth and Harold Schlosberg, *Experimental Psychology,* 2nd ed., 1954, 703.

TABLE I

PAIRED-ASSOCIATE ITEMS

MEEV-KLOF	QUIF-WHAB	HOOL-NOWK	ZERM-SUSH
TARK-ZOOD	LORZ-HICE	JISH-LAPE	THOG-BULE
KRAD-FOJE	PIBE-GREH	GLEX-TWIC	SRUT-ROIZ

TABLE II

RATIOS OF PAIRS CORRECTLY RECALLED TO PAIRS PRACTICED FOR *S*S RECEIVING NEW PAIRS ON SECOND TRIAL(S) AND *S*S REPEATING MISSED PAIRS (NS)

	Trial 1	Critical Trial 2	Non-critical Trial 2
S	15.5/120	3.5/108	1/12
NS	11/120	22/108	1/11

Results. In scoring, a credit of either 1 or 0.5 was given if the response-item was recalled. If the recall was substantially correct, a credit of 1 was given. If, as happened occasionally, *S* pronounced only one of the consonant sounds of a response-item correctly, the respose was given a credit of 0.5. The results are given in Table II. Only the number of critical paired items, *i.e.* paired items which both members of a pair of *S*s received on their second trial but which were given, and completely missed, to only the non-substitutive pair-member on the first trial, is entered in the second column. A Wilcoxin matched-pairs signed-ranks test was made. The difference of the scores of pair members on critical pairs for the second trial was significant ($p < 0.01$).

Discussion. Possibly the poorer performance of the *S*s, with whom the substitutive procedure was used, was due to proactive interference from paired items practiced and missed completely on the first trial which were then removed.[4] As an alternative explanation, the decrease from 15.5/120 to 3.5/108 may also be entirely the result of greater difficulty of the critical pairs since it was these pairs which were missed on the first trial of the non-substitutive procedures.[5] Despite any such differences in difficulty for the non-substitutive procedure, the proportion of paired items correctly recalled on the second trial, 23/119, was significantly higher than the proportion for the first trial, 11/120 ($p < 0.03$, two-tailed, *a posteriori* application of the binomial test).

On the basis of the results of this study it is concluded that the operations employed by Rock did not adequately provide for the demonstration of increasing probability of initial recall with repeated practice. Although the present study does not satisfactorily disprove Rock's thesis of one-trial learning, it does provide marginal support for the view that his results may simply be a special case of gradual initial acquisition. In testing this interpretation, further research in which proactive interference is controlled and in which such factors as exposure-rate, difficulty-level, and similarity of items are systematically varied may be of value. It may be possible either to increase or to decrease the probability of recalling items repeated missed, relative to new items, by varying any of these factors.

Summary. Twenty experimental *S*s were given one additional practice of paired associates which they missed on the first trial. The proportion of these repeated paired items which was recalled on the second trial significantly exceeded the proportion of paired items correctly recalled: (1) on the first trial, and (2) when the same paired items missed by the experimental *S*s were given to control *S*s on their second trial only. It is suggested that "one-trial" acquisition of paired associate items may be a special case of gradual initial acquisition.

---

[4] Due to an error by *E*, one pair of items which was recalled partially correctly on the first trial of an *S* with the non-substitutive procedure was retained for his second trial. This *S* recalled no critical pairs on his second testing.

[5] This interpretation is supported by the results of recent research in which each *S* received three repeated and missed paired items and three new ones for the second trial. Proactive interference could not have differentially suppressed the initial recall of new paired items as compared to repeated ones and thus account for the much larger recall of repeated items which again resulted. Unfortunately, there were several experimental errors due to the complexity of *E*'s task and a more stable substitutive procedure is needed.

APPARATUS

A MULTIPLE S–R APPARATUS FOR HUMAN LEARNING

By Arthur R. Jensen, Carter C. Collins, and Robert W. Vreeland, University of California, Berkeley

In the study of human learning, it is often desirable, when conducting a variety of experiments, to make use of a highly standardized situation which also permits the manipulation of a number of independent variables. Especially in the study of individual differences in learning, it is important that many experimental parameters be constant throughout variations in the particular independent variables under investigation. It is also essential that there be a high degree of uniformity and dependability in the procedures when many *S*s are put through the same experiment.

To meet these requirements, an electrical apparatus has been constructed which permits a wide variety of uses in experimental research on human learning.[1] It is called a multiple S–R apparatus because of its three essential features: (1) It is capable of presenting, in any desired sequence at any desired time-intervals, a variety of visual stimulus-objects (numbers, letters, words, nonsense-syllables, figures, pictures, colors, etc.); (2) it delimits *S*'s repertoire of responses by having him respond by pressing buttons which can be varied in number and in spatial arrangement; and (3) it delivers 'feedback' or 'reinforcement' on any desired schedule. The procedure for any particular experiment can be unfailingly uniform for every *S*, since the entire experimental procedure (*i.e.* the sequence of stimulus-materials and the schelule of reinforcement) is completely controlled by a program which has been coded on teletype-tape.

The apparatus has all the capabilities of the conventional memory-drum, and many more. For example, the sequence of stimulus-objects may appear in a completely random order for an indefinite number of presentations. A greater variety of objects can be presented, since they do not appear in a narrow aperture, as in the usual memory-drum, but are projected on a screen. Anything that can be photographed can be presented. *S* may respond verabally, as in the conventional memory-drum experiment, but often it is desirable clearly to define and delimit S's repertoire of responses. This the apparatus does by requiring S to respond by pressing buttons, which can be labeled as desired and varied in number and arrangement. The response-alternatives (push-buttons) are so connected with the stimulus-objects that certain S–R connections will result in 'reinforcement' (either a 'bong' or a green light). The apparatus is especially suitable for experiments on selective trial-and-error learning, schedules of reinforcement, paired-associates learning, serial learning, discriminative learning, recognition of symbols, probability-matching, concept-attainment, simple problem-solving, and, in general, for any type of experiment in which S must learn multiple responses to multiple stimulus objects.

---

[1] The design and construction of this apparatus was made possible by a grant-in-aid by the Research Fund Committee of the School of Education.

The apparatus consists of four units: (1) the teletype-tape distributor, (2) the stimulus-display unit, (3) S's response-panel, and (4) E's control-console. S is alone in one room with only the stimulus-display unit and the response-panel, while E is in an adjoining room (with an intercom and one-way vision window) containing the teletype-machine and the control-console. The parts of the apparatus to which S is exposed are completely silent in their operation, there being no mechanical parts. S is thus completely free from any environmental distractions and may concentrate solely on the stimulus-objects and responses essential to the experimental task.

A primary concern in the design of the apparatus was high reliability of performance. This has been achieved. The relays, the teletype 'reader,' and the stimulus-display unit all are of the highest quality commercially available; they were, in fact, perfected for durability and dependability in uses involving much greater stress and strain than would ever be demanded of the present apparatus.

DESCRIPTION OF THE APPARATUS

Teletype-tape distributor. Each stimulus-object presented by the stimulus-display unit is coded on a five-hole teletype-tape. The sequence of stimulus-presentations for the entire experiment is punched on the tape. Reinforcement for the 'correct' response is also programmed on the tape, hence any desired schedule of intermittent reinforcement is possible. The tape holes are read sequentially by a standard Western Union teletype-distributor, shown in Fig. 1. A 4-relay tree provides a 16-character readout code. Holding contacts retain the coded character for a duration that is set on E's control-console.

Stimulus-display unit. This unit, shown in Fig. 2, consists essentially of an In-Line Digital Display unit (Series 80,000) manufactured by Industrial Electronic Engineers, Inc. It has been modified by cutting a slot in the side of it so that the entire set of stimulus-objects may be readily changed by removing the plate of 12 condensing lenses to which the objects are attached.

The objects appear singly on a solid, black background at the center of the 3¼ × 5¼ in. display-screen. They are sharp and clear and, if they are numbers, letters, or words, may usually be read from a distance of up to 100 ft. (The object being presented in Fig. 2 is a blue triangle.)

Fig. 3 shows the construction of the stimulus-display unit. It is actually a set of

12 miniature projectors, each of which can be activated, via the teletype-tape distributor, either independently or in combination with the other projectors. Any object that can be photographed can be presented; negatives of the photographically reduced images are fastened to the plate of condensing lenses (G in Fig. 3). Either colored or black-and white images are possible. The projectors are illuminated by twelve 24-v. lamps (II, 12, etc. in Fig. 6). Diodoes (D1 and D2 in Fig. 6) permit the negative 24-v. power from the relay-tree to light the lamps but block the 175-v. positive 'error bias.' The duration of stimulus-display, which can be set on *E*'s console, are controlled by a 6D4 vacuum-tube relaxation-oscillator which closes a relay in the 'pacing interval-timer' (Fig. 6). This relay also actuates the 'total stimuli' counter of *E*'s console and advances the teletype-tape.

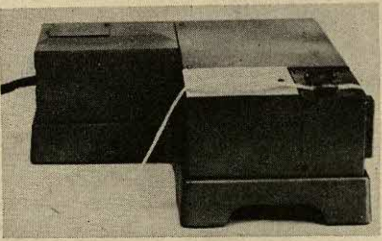


FIG. 1. WESTERN UNION TELETYPE-TAPE DISTRIBUTOR

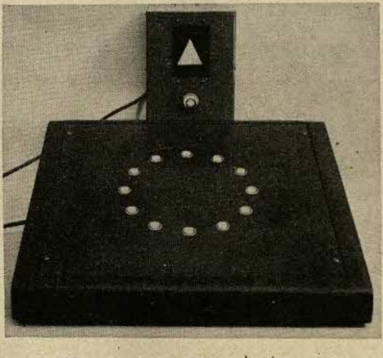


FIG. 2. THE DISPLAY-UNIT

Directly below the display-screen is the signal of reinforcement, a green light which goes on the instant *S* makes a 'correct' response. With machine-pacing, the light comes on when the button is pressed and stays on until the end of the machine-pacing interval (*i.e.* 2–10 sec.). Under subject-pacing, the reinforcement-light stays on 1½ sec. regardless of the duration of *S*'s response. A single doorbell 'bong' is contained in the display-unit, and the sound of the 'bong' may serve as the reinforcer instead of the light. The mode of reinforcement (*i.e.* visual or auditory) is controlled by a switch on *E*'s control-console.

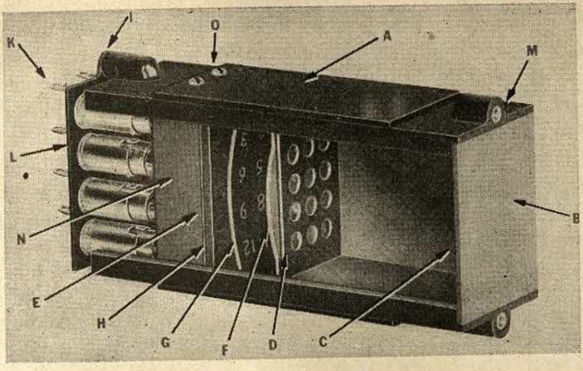


FIG. 3. INTERIOR OF THE DISPLAY-UNIT

(A) Heat-radiator—aluminum alloy. (B) Viewing screen, lucite. (C) Diffusion coating on inner side. (D) Mask to stop stray light. (E) Gasket—commercial grade, gray vulcanized fiber. (F) Projection-lens, lucite. (G) Condensing lens, lucite; stimulus-objects on concave surface. (H) Heat-absorbing glass to remove infra-red rays. (I) Spacer for ventilation of lamp. (K) Stimulus-object terminal (12 available). (L) Socket-assembly. (M) Case—Dow Chemical #440, high-impact, high-heat resistant polystyrene. (N) Metal light-shield and heat-sink. (O) Aluminum screw connects heat-sink to radiator.

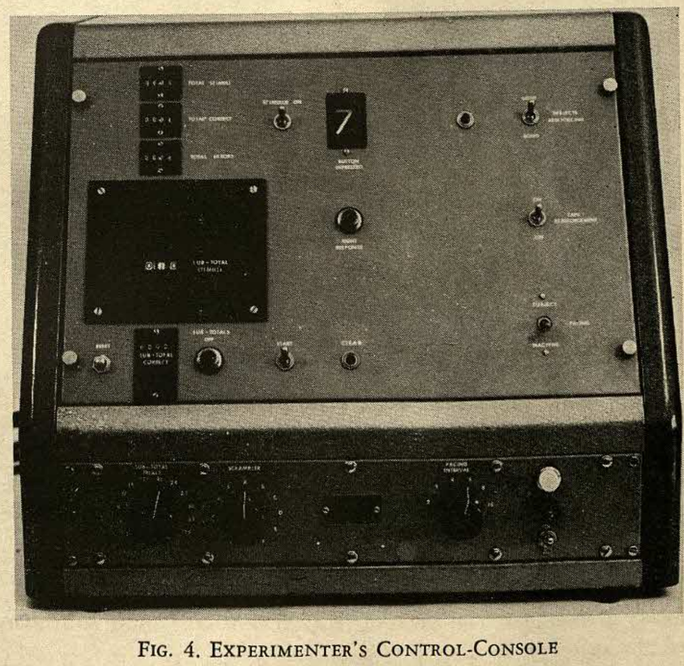


FIG. 4. EXPERIMENTER'S CONTROL-CONSOLE

S's response-panel. *S* responds to each presentation of a stimulus by depressing one button in an array of push-buttons. A circular array of 12 buttons is shown in Fig. 2, but the panel into which the push-buttons are plugged can readily be removed and replaced by other panels having different arrays and different numbers of buttons. Each button can be plugged into a panel resembling a PBX telephone-switchboard underneath the response-panel. The push-buttons are flat-topped that labels may be easily attached. When *S* presses the 'correct' button for a given stimulus, there is immediate reinforcement (either a green light or a 'bong' but only if the program calls for reinforcement). If *S* presses a 'wrong' button for a given stimulus, there is of course no reinforcement, and all other buttons go 'dead,' *i.e.* there are 12 holding relays (K1, Fig. 6, for example) which prevent the activation of more than a single button after any one presentation of a stimulus. Thus, if *S*

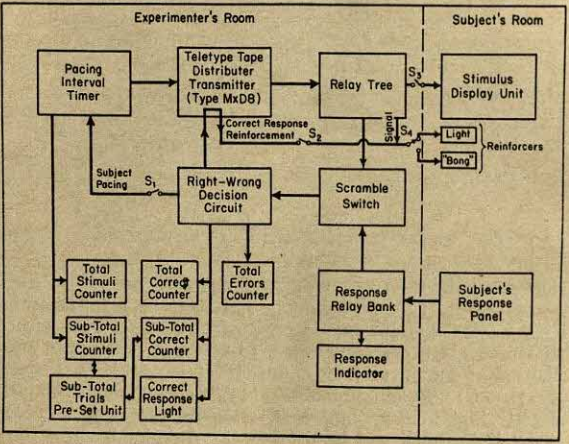


FIG. 5. MAIN COMPONENTS OF THE APPARATUS

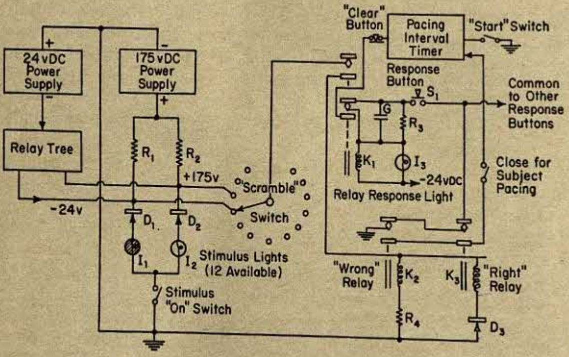


FIG. 6. THE MULTIPLE S–R CIRCUIT

makes an incorrect response, he cannot correct it and receive reinforcement on the very same trial. For example, suppose that a button (*e.g.* S1 in Fig. 6) is depressed; this completes a circuit from the —24-v. supply through the coil of K1 and contacts of K1, K3, and K2 to ground; the relay K1 closes and breaks the circuit from its coil to S1. At the same time K1 picks up a holding circuit to ground via the 'clear' button (on control-console), the 'pacing interval-time,' and the 'start' switch. Relay K1 remains closed until this holding circuit is broken. When K1 closes, it completes the circuit from the —24-v. stimulus supply throughout the relay-tree and the scramble-switch to the 'right' and 'wrong' relays. The 'right' relay closes and breaks the ground-return of all response-buttons, thus preventing the possibility of recording a second response during the presentation of the stimulus. When the machine is set for subject-pacing, the 'right' relay grounds the 6D4 cathode through the 'subject-pacing' switch, thereby advancing the tape. If *S*'s response is wrong, the 'wrong' relay closes; this relay will not close on less than about 35 v. Diode D3 prevents the 'right' relay from closing with positive potentials. All the circuits of the stimulus-lamps, except the one energized by the relay-tree, are biased by the +175-v. 'wrong' bias-supply.

E's control-console. *E*'s console, shown in Fig. 4, has the following features.

(1) Power-switch. The main power-switch turns on the entire apparatus. A pink bulb over the switch is the pilot-light.

(2) Pacing-timer. The pacing-timer controls the rate of presentation of stimuli, which can vary from 2 to 10 sec. per stimulus. A sequence of stimuli separated by 'blank' intervals may be programed on the teletype-tape.

(3) Pacing-switch. The pacing switch can be put on machine-pacing or *S*-pacing. When the switch is set on *S*-pacing, the stimulus does not change until *S* makes a response (*i.e.* presses a button). The stimulus remains on the display-screen for 1½ sec. after *S* has responded (whether 'correctly' or 'incorrectly'), thereby permitting the reinforcement to appear (or to be heard) while the stimulus is still in view.

(4) Tape-switch. The tape-switch turns off the reinforcement even though it is programed on the tape. Thus *E* can stop reinforcement at any time in the course of the experiment.

(5) S's switch. *S*'s switch determines whether the reinforcement shall be auditory (a 'bong') or visual (a green light below the stimulus-display screen). A button to the left of the *S*'s reinforcing switch permits *E* to deliver reinforcement that is not programed on the tape.

(6) Scrambler. The scrambler-dial changes the connections between the stimuli and the push-buttons on the response-panel. There are 12 sets of S–R connections, all of which are random in relation to one another, so that *S* cannot systematically 'transpose' from one set of connections to another. Thus there can be 12 'equivalent forms' of any learning task programed on a single tape.

(7) Stimulus-switch. The stimulus-switch turns the lamps in the display-unit 'off' and 'on.' It permits *E* to eliminate the stimuli while the tape continues to advance through the distributor.

(8) Start-stop switch. This switch starts and stops the movement of the tape through the teletype-tape distributor.

(9) Right response. The correct response is indicated by a green light which turns on when *S* makes a correct response, whether or not he received reinforcement.

(10) Buttons depressed. The button-depressed screen consists of a 1⁹⁄₁₆ × 2⅝ in. In-Line Digital Display Unit (Series 10,000, manufactured by Industrial Electronic Engineers, Inc.) which indicates the number of the response-button *S* has pressed. The number remains in view for 1½ sec. after *S* has pressed a button or for the remainder of the machine-pacing interval, permitting *E* to record *S*'s responses throughout the experiment. In the present apparatus there is no automatic permanent recording of *S*'s responses. This feature could easily be made possible by attaching a

recording device to the same circuits that activate the response-indicator ('Button Depressed') and the 'Right Response' light in the control-console.

(11) Electrical counters. There are five electrical counters, each with a reset switch. (a) One gives the number of correct responses (whether reinforcement or not) made by *S* throughout the experiment. (b) A second gives the total number of stimulus-presentations throughout the experiment. (c) A third gives the number of incorrect responses made by *S*. His failures to respond = Total stimulus-presentations − (Total correct responses + Total errors). (d) and (e) The fourth and fifth counters give, respectively, sub-totals for the number of stimulus-presentations and the number of correct responses made by *S*. Both of these counters are connected with the Sub-Total Dial and they serve primarily as criterial counters. The Sub-total Trials counter is set for any number of trials (*i.e.* stimulus-presentations) from 3 to 36. Say, for example, it has been set for 12. Then, after 12 stimuli have been presented, the Sub-total Stimuli counter will register 12 and the Sub-total Correct counter will indicate the number of correct responses *S* made to the last 12 stimuli. At this instant, a red indicator lights up Sub-totals Off (see Fig. 4). *E* can record the Sub-total Correct and press the Reset button, which instantly resets all the sub-total counters back to zero. If the criterion in this experiment is 12 consecutive correct responses, *E* will turn off the apparatus with the main power-switch the moment both sub-total counters register 12, which is signaled by the flashing on of the red light.

A TEST OF THE ABILITY TO LIP-READ—USE VISIBLE CUES OF SPEECH

By THOMAS FARRIMOND, Liverpool, England

The test described here was prepared for the purpose of assessing the ability of people to read 'lips'—to understand the visual components of speech as conditioned by age and loss of hearing. From the reports of others who have investigated the ability of lip-reading under conditions of noise, it was clear that, when speech is masked by noise, greater use is made of the visual patterns associated with speech. Since suitable tests are not available for measuring a 'normal' adult, the present one was constructed.

Rationale. For a complete standardization of test-conditions, a ciné-film was thought to be preferable to a 'live-voice.' Although a film may be considered artificial in some respects, it does insure that all the *S*s viewing it are given precisely the same information. This is not possible when the live-voice, even though monitored, is employed. Since, under present-day conditions, almost everyone is familiar with cinema and television, each being a two-dimensional form of presentation, it was not expected that any difficulties would be encountered by using a film. This was confirmed when the test was subsequently completed and administered to the *S*s.

The choice of material suitable for inclusion in the film was influenced by several considerations. Since *S*s of all ages and social status were to be tested, the items selected were relatively free from the influences of different educational or socioeconomic conditions. The most suitable material appeared to be (1) lists of numbers and (2) simple sentences.

Numbers-test. The first part of the film consists of a series of 50 numbers spoken by a young woman with 6-sec. pauses between the successive numbers. In this task, the *S*s had no contextual guidance and scores obtained in the test depended solely on the accuracy with which the visual components of speech were perceived.

A pilot-study using 10 *S*s of both sexes, aged from 17 to 42 yr., was conducted and an item-analysis of the results was made. From the information gained, it was possible to arrange the difficulty-level of items in the final test in such a way that there were no long periods in which *S*s were unable to give a response. Table I shows some of the numbers used during the pilot study and the percentage of times these were identified correctly by the 10 *S*s. For example, No. 4 was identified by all the *S*s thus giving it a score of 100%; No. 53 was identified correctly by only half the *S*s and was therefore scored 50%. To extend the range, three-digit numbers of known level of difficulty were included in the final version of the test.

---

* This test was prepared at the Medical Research Council's Group for Research on the Occupational Aspects of Aging, Liverpool. The author acknowledges the assistance of Mr. J. E. Fisher of the Department of Psychology, University of Liverpool.

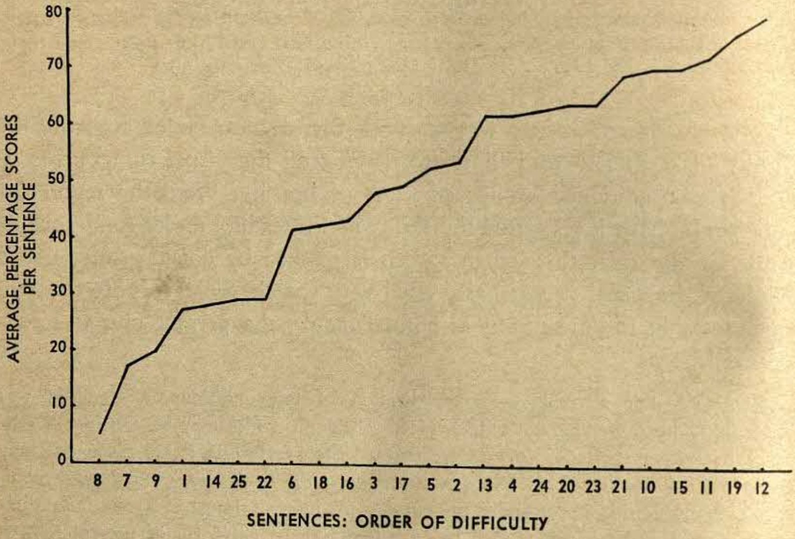


TABLE I

PERCENTAGE OF TIMES THE VARIOUS NUMBERS USED WERE CORRECTLY IDENTIFIED

% correct	Numbers used	% correct	Numbers used	% correct	Numbers used
0	3, 13, 30, 70	40	14, 44, 69, 85	70	6, 18, 31, 54, 62, 81
10	87	50	28, 40, 53	80	24, 47, 82
20	5, 53, 60	60	1, 8, 11, 45, 57	90	12, 25, 37, 72
30	15, 23, 38, 58, 88, 96			100	2, 4, 20, 21

TABLE II

NUMBERS GROUPED INTO SIMILAR SAMPLES

Groups	Numbers in order of presentation
Examples:	24, 32, 124, 375, 78
Form A:	20, 2, 37, 123, 72, 82, 47, 246, 6, 81, 57, 8, 45, 1, 179, 28, 87, 205, 44, 88, 53, 198, 13, 500, 22
Form B:	21, 4, 12, 234, 25, 24, 62, 375, 54, 16, 89, 14, 11, 53, 387, 40, 108, 85, 38, 23, 3, 297, 45, 160, 27

The illustrative examples and the 50 numbers that were finally selected for the test, which was divided into two similar and equally difficult forms (Forms A and B), are shown in Table II. Form A commences with '20' and ends with '22,' respectively; both easily lip-read. Throughout the test, easy 'numbers' were interposed among difficult ones that *S*'s motivation would be sustained—that he would not become discouraged and give up. Form B, similar in construction to Form A, was given immediately after Form A, but it may be given after any interval desired.

Two points were scored for every number correctly identified by *S*, hence the scores obtained, since 50 numbers were used, could be translated directly into percentages. Table III shows the numbers used in the final version of the test and their difficulty as determined from an item-analysis of the scores of 179 *S*s. The product-moment correlation between the scores on Forms A and B is + 0.70.

The projected image of the speaker's head, face, and shoulders was life-size. It was shown at a visual contrast approximating average, daylight conditions. The person selected to read the numbers spoke 'standard' English and she had had some experience in dramatics. She repeated the numbers in a normal conversational manner without exaggerating the pronunciation. Every 'exposure' in the series, *i.e.* the pronunciation of every number, was marked at the bottom left corner of the film that, if so desired, the test could be given to groups instead of individually. This 'mark' could easily be eliminated from the screen by interposing a card at the appropriate position in the beam of the projector.

TABLE III

NUMBERS USED IN THE FINAL VERSION OF FORMS A AND B

% Correct	Form A	Form B
0–10	13, 47, 198, 205, 246, 500	16, 24, 108, 160, 234, 297, 375, 387
11–20	1, 6, 44, 53, 123, 179	3, 14, 38, 40, 85
21–30	57, 87	53, 54, 62
31–40	37, 45, 81, 88	89
41–50	28, 82	12, 25, 45
51–60	8, 20	11
61–70	72	23, 27
71–80		4, 21
81–90	22	
91–100	2	

Sentences. The second part of the film was constructed along similar lines, using sentences in place of numbers. In this part, some contextual guidance was given in the form of a relevant scene preceding every spoken sentence. Many different scenes were recorded; simply every-day activities such as, for example, a car crossing a bridge, a man driving a nail into a piece of wood with a hammer, etc. From among these scenes, 25 were chosen, each of approximately 6 sec., and a picture of a woman speaking a short sentence was spliced after each. The list of the 25 sentences finally selected is shown in Table IV.

The italicized words are the scorable words. If they were correctly identified, irrespective of the order in which they were reported, *S* received one credit for each word. Plurals and changes in tense were, however, counted as incorrect. Among the 25 sentences chosen, there are 100 scorable words. Opposite the sentences in Table

IV, the rank order of their difficulty, from most to least difficult, is shown. Sentence 8 is the most difficult; Sentence 12 is the least difficult.

In the film, the scenes and their accompanying sentences were so arranged (see Table IV) that the proportion of easy to difficult sentences was the same in both halves of the list. This arrangement was also chosen, as in the case of the numbers, that the *Ss* would be able to respond to some items throughout the test-period, and hence would not lose interest in the task.

The degree of relevance of the sentence to its previous scene was adjusted to make the test discriminate among the *Ss* who guessed on the basis of the context and those who followed the lip-movements more accurately. For example, after showing a scene in which bread was being thrown to sparrows and ducks, the speaker (Sentence No. 9), said: "The ducks eat the bread." Typical responses to this sentence range from "The birds are hungry" to an accurate word-for-word reproduction of what was said. Other items in the series were not capable of a very wide interpretation because the activity was simple and the amount of visual information available was small. One such example is that of a man hammering a nail into a piece of wood. The visual information is restricted to the important objects concerned in the activity, the hammer, the wood and the nail—little irrelevant

TABLE IV

LIST OF SENTENCES SPOKEN

Sentences (spoken in order)	Scorable words	Order of difficulty
(1) The *drill goes into* the *wood*	4	4
(2) The *hammer hits* the *nail*	3	14
(3) He *tied* the *shoe lace quickly*	4	11
(4) The *man sharpened his pencil*	4	16
(5) The *man lit his cigarette*	4	13
(6) The *tea was poured into* the *cup*	5	8
(7) The *ducks swim to* the *bank*	4	2
(8) The *angler makes* a *cast*	3	1
(9) The *ducks eat* the *bread*	3	3
(10) The *man sat on* the *seat*	4	21
(11) The *woman pushed* the *pram*	3	23
(12) The *baby sat in* the *pram*	4	25
(13) The *car went across* a *bridge*	4	15
(14) *Black smoke came from* the *funnel*	5	5
(15) The *water was poured into* a *pail*	5	22
(16) He *wound up his watch*	4	10
(17) The *paper was cut in half*	5	12
(18) A *knot was tied in* the *string*	5	9
(19) The *man dug* the *garden*	3	24
(20) He *pumped up* the *tire*	3	18
(21) The *girl combed her hair*	4	20
(22) *She took* a *purse from her handbag*	6	7
(23) The *man took* a *photograph*	3	19
(24) *She brushed up* the *broken glass*	5	17
(25) The *girl wrote* a *letter*	3	6
Total	100	

or misleading information is present that the responses obtained on this item are more accurate and show less variation.

As used by the writer, all responses required of the *Ss*, in both the number- and the sentence-tests, were purely verbal, no writing was called for at any point. Group administration is possible, however, if desired and scoring sheets were prepared for this purpose. The viewing distance for all *Ss* was 5 ft., and the viewing angle relative to the normal at the center of the white matt screen was approximately 20°.

A white calibration square one square foot in area is included before the examples on the film that intensity of illumination and size of image may be adjusted. Through the test-period a 15-w. electric lamp was arranged to give a low-level overall illumination. This lamp was above and slightly behind the screen and well out of *S*'s line of vision. The illumination provided by the lamp was just sufficient to enable *E* to enter on a record sheet the responses of *S*.

After five items for practice, few of the *Ss* experienced any difficulty in understanding what was required of them. Since all responses were individually scored by *E*, any tendency on the part of certain *Ss* to give stereotyped responses without reference to the film was quickly noticed and prevented. Indeed, it occurred but once during the practice-period and only twice during the test-proper.

The tests, both number and sentence, were given to 179 *Ss* drawn from industry. The *Ss*' ages, which averaged 43.1 yr., *SD* 11.3 yr., ranged from 20 to 79 yr. The product-moment correlations between the tests is +0.74. The individual scores obtained with sentences averaged 48.8%, *SD* 17.0%, and ranged from 5% to 92%; and with numbers, averaged 31.2%, *SD* 14.1%, and ranged from 0% to 68%.

No marked individual differences were observed in the ability of our *Ss* to interpret accurately the visual cues presented in the two tests. Though age and vocabulary played a small part in determining performance, hearing losses of the order encountered among our *Ss* did not play an important role.[1] The sentence-test has, however, been very useful in obtaining data on visual recall and, in conjunction with the tape-recordings of the *Ss*' ability to hear, in making comparisons between visual and auditory recall.

---

[1] Thomas Farrimond, Age-differences in the ability to use visual cues in auditory communication, *Language and Speech*, 2, 1959, 179-192.

APPARATUS AND PROCEDURES FOR STUDYING TASTE-PREFERENCES IN THE WHITE RAT

By PAUL THOMAS YOUNG and WILLIAM E. KAPPAUF,
University of Illinois

The purpose of this paper is to describe apparatus and procedures currently used by us in experiments on taste-preferences of the white rat. The apparatus was developed for use with liquid foods or substances in solution. Two solutions, presented in a traditional two-choice drinking situation, are simultaneously available to *S* for a period of 3–4 min., and *E* determines which of the two solutions *S* licks more frequently. Advantages seen for the present technique in comparison with other methods of preference-testing include the use of a forced-sampling sequence prior to the preference-test, the method of quantifying *S*'s contact with each solution, and the general speed with which preference-studies can be conducted.

Apparatus. The unit into which *S* is placed for testing is shown in Fig. 1. It is a box, 5 in. wide, 10 in. long, and 6.5 in. deep (inside dimensions). The forward wall of the box is of clear plastic, and it is through apertures in this wall that two drinking tubes are inserted. The remaining walls are of wood. The hinged cover and rear portion of the floor are of ½-in. hardware cloth. The forward portion of the floor is of plastic. The drinking tubes are separated 1¼ in. laterally and are 2¼ in. above the level of the floor. They extend approximately ⅛ in. into the box when testing is in progress.

Drinking can be monitored in any of several ways. In some series, *E* has observed *S* directly through the plastic wall and has operated a three-position, double-throw switch to accumulate on separate stop-clocks the time of drinking at each tube. In most tests, however, drinking from each tube has been recorded electrically through a 'drinkometer' circuit operating a counter. At one time these circuits were used to operate cumulative recorders. Because the test-box is not itself enclosed and sound-proofed, appropriate steps have been taken to screen *E* and to quiet the timing and recording devices.

The drinkometer circuit which has been used thus far is that described by Hunter.[1] With each lick at one of the solutions, *S* triggers the recorder by completing the local circuit between the solution and the wire portion of the floor with which its hind feet are in contact. Isolation of the forward wall of the test-box from the floor and the use of plastic for the front part of the floor prevents the circuit from being shorted as a result of accidental dripping of the solution. *S*'s forepaws, which sometimes are wet from direct contact with the tubes, rest on the plastic portion of the floor and hence are not involved in the circuit.

---

* The development of this apparatus was supported in part by the Research Board of the University of Illinois, and in part by Research Grant G-14538 from the National Science Foundation.

[1] T. A. Hunter, An electronic contact relay, this JOURNAL, 65, 1952, 458-459.

The current which flows through *S* when contact is made with the solution has been found to be about 2 μa. At this level, the degree of ionization of the solution as it reaches *S*'s tongue is exceedingly low and is believed not to contribute to *S*'s differential response to the two solutions. The possibility that the current might have a direct stimulating effect on the taste-receptors of the rat is discounted in view of the fact that it is below the human detection-threshold.

The drinking tubes are of glass, shaped like those used in laboratory water-bottles. The two tubes on a single box are matched for size of the nozzle-openings. Into the heel or bend of each tube, a fine platinum wire is inserted and sealed in place. This wire extends approximately 2 mm. into the test-solution. The wire is connected with a solder-lug to which an alligator-clip may be attached as shown in Fig. 1. The negative side of the recording circuit is led to this wire, the positive side to the cage-floor.

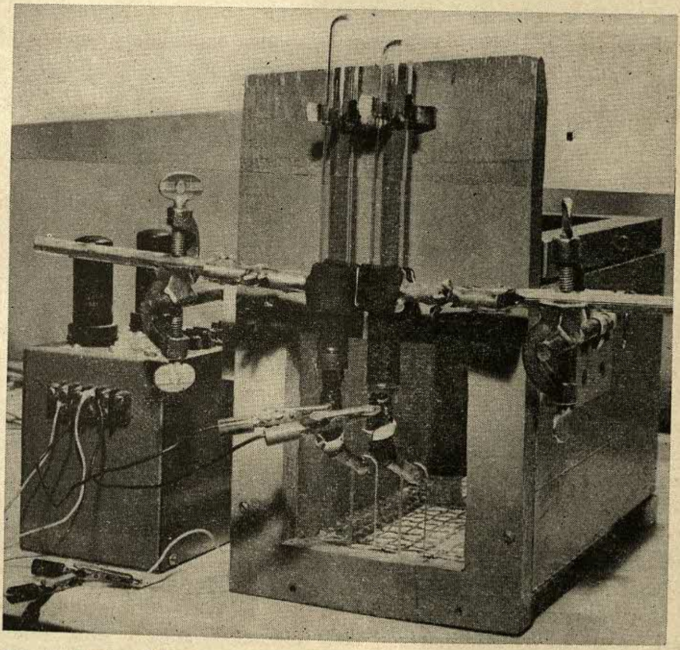


FIG. 1. FRONT VIEW OF THE TESTING BOX FOR TASTE-PREFERENCES
(A central vertical portion of this photograph, showing only the full length of the drinking tubes, has been deleted.)

The burettes containing the test-solutions are attached to the test-box in the manner shown. Each burette is supported by an upper clip and positioned at the base by an apparatus-clamp. Partial release of the apparatus-clamp allows the drinking tube and lower portion of the burette to be swung in and out about the upper clip as a pivot. This permits the rapid removal of either or both drinking tubes as well as the adjustment of the extent of insertion of the tubes into the box.

The equipment presently in use consists of six testing boxes arranged side by side. The leads from the six units pass to separate amplification-units, at the left of the testing boxes, and thence to relays and to the recording mechanisms.[2]

Preliminary training. Animals are trained with a simple sucrose solution from 8 to 10% (weight/volume) in both drinking tubes. With first one tube and then the other available, the *S*s learn to drink from either nozzle. Then they are offered a choice between two solutions which are known to be readily discriminable and between which a preference exists for all animals, *e.g.*, 9% sucrose vs. distilled water.[3] Every *S* is placed in the apparatus daily for a short time until, regardless of right-left position of individual solutions, it exhibits a consistent preference. This preliminary training usually requires from 5 to 10 days. *S* then is ready for testing in situations involving other solutions.

Forced-sampling procedure. As testing with other solutions begins, *S* is forced to sample each solution once or twice before being offered the pair of drinking tubes for the preference-test. The procedure is as follows: Neither tube is in place when *S* first is put in the test-box. When *S* comes to the front of the box, one tube is inserted and remains there until about five licks have been taken. Then that tube is withdrawn and the other inserted for some five additional licks. This sequence may be repeated for the first tube and for the second. Shortly after the second tube is withdrawn, at the end of sampling, both tubes are inserted simultaneously and the testing interval begins. The *S*s adapt well to this routine.

Exposure-time in preference-trials. The major purpose for which the apparatus is being used is the study of *S*'s initial taste-preferences—its preferences as it begins to drink. For these studies, the *S*s are nondeprived and free from all known metabolic need, being maintained *ad libitum* on a diet which is nutritionally complete. *E* relies on the incentive-character of the solutions to elicit and sustain drinking.

Concern for initial taste-preferences implies that the drinking or exposure-time should be short. Indeed, in previous work with a different apparatus and method,[4] exposure-times of a second or two and sometimes less than a second have been used, since the quality of choice was the main criterion of preference. In the present work, however, with a criterion of preference based on the relative frequency of tongue-contacts, it was clear from the start that a compromise would have to be made between very brief exposure-times which might provide unreliable observations and longer times which might involve the loss of information about initial preferences. Considerable data therefore were collected relating to this matter. They have led to the adoption of exposure-intervals of 3-4 min.

Psychophysical procedures. The testing of an *S* on one day by the procedure described above constitutes one trial in a series of trials conducted according to some psychophysical routine. Observations may be compiled across days for the same *S*, or across animals for a given day. Thus far, preference-thresholds have been measured by the method of constant stimuli; aversion-thresholds have been measured by the method of limits; and pairs of equally acceptable solutions have been determined either by the method of limits or by the up-and-down method. The use of the last method has been particularly successful. It makes possible the reliable determination of a pair of equally acceptable solutions in a single day's testing with a group of *S*s.

---

[2] Counters are presently in use. Veeder-Root counters are recommended.

[3] P. T. Young and J. T. Greene, Quantity of food ingested as a measure of relative acceptability, *J. comp. physiol. Psychol.*, 46, 1953, 288-294.

[4] An early form of the brief-exposure preference-tester is described by Young (Running activity and dietary habit of the rat in relation to food preference, *J. comp. Psychol.*, 37, 1944, 327-370). A later form, substituting cups for food tubes, is pictured in C. P. Stone (ed.), *Comparative Psychology*, 3rd ed., 1951, 73-74.

NOTES AND DISCUSSIONS

THE STAIRCASE-METHOD IN PSYCHOPHYSICS

A psychophysical method variously referred to as the "method of up and downs,"[1] the Békésy audiometric method,[2] or the staircase-method, has come into extensive use in the last few years. The method has several advantages over other more commonly used techniques but it also has some disadvantages. This paper will illustrate the use of the method, will discuss its relative merits and demerits, and will describe a modification which overcomes certain of the disadvantages of the method.

The staircase-method is best described by illustrating its use with a specific problem. Suppose the problem is to determine *S*'s absolute, intensive threshold for the sound of a click. The first stimulus that *E* delivers is a click of some arbitrary intensity. *S* responds either that he did or did not hear it. If *S* says 'yes' (he did hear it), the next stimulus is made less intense, and if *S* says 'no,' the second stimulus is made more intense. If *S* responds 'yes' to the second stimulus, the third is made less intense, and if he says 'no,' it is made more intense. This procedure is simply continued until some predetermined criterion or 'number of trials' is reached. The results of a series of 30 trials are shown in Fig. 1. The results may be recorded directly on graph-paper; doing so helps *E* keep the procedure straight.

There are a number of ways of determining the intensive value that represents the threshold. The simplest is to compute the mean of the values of a given number of stimuli delivered after the series has reached its final level. This requires an arbitrary decision about when the final level has been reached. The technique, which avoids this difficulty and yields a 50% value, is simply to determine the stimulus above which 50% of the responses are 'yes,'—*i.e.* in Fig. 1 between 61 and 62 db.

Statistical treatment of the results has been discussed by Dixon and Massey, who describe the techniques for determining the means, standard deviations, standard errors, etc., for this type of data.[3] The treatments assume, however, that the response to each stimulus is independent of the preceding stimuli and preceding responses. This assumption holds for the examples analyzed, but there is evidence that the assumption does not always hold for human *Ss* in psychophysical experiments.[4] The development of techniques that take the existing inter-actions into account has not as yet been achieved.

---

[1] W. J. Dixon and F. J. Massey, *Introduction to Statistical Analysis,* 1957, 279-286.

[2] Georg von Békésy, A new audiometer, *Acta Oto-laryngol.,* 35, 1947, 411-422.

[3] Dixon and Massey, *op. cit.,* 286.

[4] W. S. Verplanck, G. H. Collier, and J. W. Cotton, Nonindependence of successive responses in measurement of the visual threshold, *J. exp. Psychol.,* 42, 1952, 273-282; Verplanck and Cotton, The dependence of frequencies of seeing on procedural variables: I. Direction and length of series of intensity-ordered stimuli, *J. gen. Psychol.,* 53, 1955, 37-47; V. L. Senders, Further analysis of response sequences in the setting of a psychophysical experiment, this JOURNAL, 66, 1953, 215-229; R. S. Woodworth and Harold Schlosberg, *Experimental Psychology,* 1954, 225-233.

An important characteristic of the staircase-method is that *E* imposes four predetermined conditions upon every series. He decides in advance (1) where to start the series; (2) how large the steps are; (3) when the series should be stopped; and (4) when the series should be modified.

(1) Where to start. As in most psychophysical methods, several preliminary series are necessary to establish the range in which the threshold falls. In general the method is most efficient if the first stimulus is so near the threshold-level that many trials are not required in approaching the threshold.

(2) Step-size. Stimulus-intensities should be chosen along some scale that yields approximately equal sensory intervals. Logarithmic steps suffice for most applications of the method. The size of the steps should be so selected that no more than two, three, or four like responses are made before a reversal of the responses occurs.

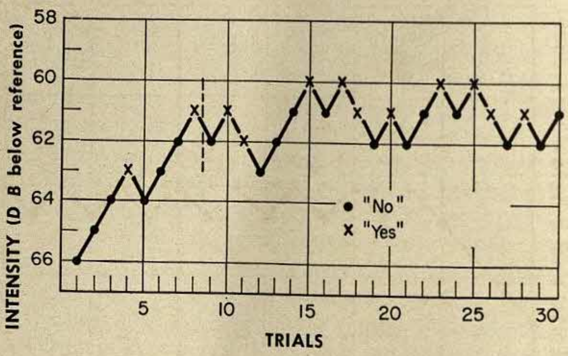


FIG. 1. DATA FROM THE DETERMINATION OF A TYPICAL AUDITORY THRESHOLD BY THE STAIRCASE-METHOD

For instance, if the step is very large, *S* will simply alternate 'yes' and 'no' as the intensity of the stimulus jumps from above to below his threshold. In this case, the value of the threshold would be insensitive to changes in the experimental variables. If, on the other hand, the step-size is very small, *S* will give long series of 'yes' or 'no.' When this happens, the method is inefficient because the succeeding stimulus-responses do not give as much information as they would if the step-size were larger. That is, if *S* has already responded with a 'no' (or 'yes') to a given intensity, the fact that he says 'no' (or 'yes') to an intensity only very slightly greater (or smaller) could have been predicted with a large degree of certainty. Therefore, when he actually does say 'no,' (or 'yes') not much new information is received. In general, this psychophysical method, as well as all others, becomes maximally efficient when the stimulus-steps are the size of the differential threshold.[5] Should this information be available, preliminary experimenting with step-size is unnecessary.

(3) When to stop. In general, the results with the staircase-method are like those shown in Fig. 1. The values of the stimuli presented change relatively rapidly un-

---

[5] Dixon and Massey, *op. cit.*, 289.

til they reach an asymptotic level or plateau, and then they hover around this level as long as the conditions remain unchanged. Obviously, the longer the series, the more reliable will be the computed value of the threshold. Thus, the decision about when to end the series is a compromise between a large number of trials for reliability and a small number for economy in time. The particular compromise depends upon the particular requirements of the experiment. The simplest method of determining when to end the series is to decide to conduct a predetermined number of trials. There are more efficient procedures, however, which take account of the fact that the values of the first several trials rest to a large degree upon where *E* began the series. These trials should not, therefore, be included in the final computation of the threshold. A better method is to decide, in advance, to end the series at a predetermined number of stimulus-responses after the plateau has been reached, and to include only these latter responses in the computation of the threshold-value. The problem, then, is to decide when the final level has been reached.

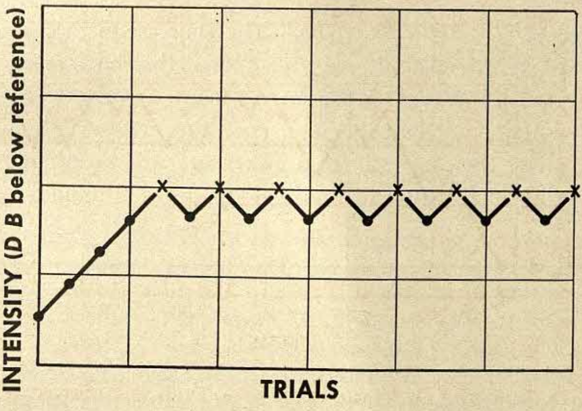


FIG. 2. DATA FROM A THRESHOLD-DETERMINATION UNDER IDEALIZED CONDITIONS

If there were no variability in the apparatus or in *S*, and if every response was completely independent of the previous ones, the data would look like those shown in Fig. 2 and *E* could simply say that the final level was reached at the intensity where the first reversal of response occurred. Data from human *Ss*, however, show evidence both of variability and of interdependence of responses and a plot of the results usually obtained is like that shown in Fig. 1. To allow for variability in the apparatus and also in the *S*, we may decide that the final level has been reached after some arbitrary number of reversals greater than one has been reached. The particular number of reversals chosen depends upon the amount of variability and interdependency that occur under the particular conditions of the experiment, and this must be discovered in preliminary trials. For example, if the data in Fig. 1 were the result of preliminary trials, we should decide that the experimental run will consist of 10 stimuli delivered after the third reversal—after the dashed vertical line in Fig. 1.

(4) Modification of step-sizes. Under some conditions, the size of the steps in

staircase-method should be varied during the course of the experiment. For example, if the method were used to determine the absolute, visual, intensive threshold, the steps in the preliminary stage of dark-adaptation should be large and in the final stage they should be much smaller. From the point of view of careful experimental design, it would be desirable that the first stimulus in each of the staircases be at the same intensity-level. As a result, in at least one of the series, and probably in both, the starting level would be far removed from the final level. This may be avoided by using large steps in getting to the final level and then switching to smaller steps when the final level has been reached. For example, it may be decided in advance that, for the visual-threshold, the steps will be 50% changes in intensity until the third reversal, and 10% changes thereafter. As a rule of thumb, when this sort of procedure is used, a series that contains a total of 20–25 trials will produce a fairly reliable threshold-value under a number of diverse conditions.

Advantages of the staircase-method. The staircase-method is extremely efficient. For a given reliability of a computed threshold-value, the staircase-method requires the presentation of many fewer stimuli than any other psychophysical method, because, once the first few stimuli are out of the way, all of the other stimuli are very near the threshold-level, each one contributing importantly to the final computed threshold-value.

Comparison with method of adjustment. A comparison of the staircase-method with the method of adjustment is difficult. The two methods are very similar, since, in the method of adjustment, *S* changes the stimuli in much the same way that *E* does in the staircase-method. To the extent that this is true, the two methods may be equally efficient. The staircase-method has, however, one advantage in that there is no ambiguity about what is actually happening. With the method of adjustment, the particular values of the stimuli that *S* has experienced are usually unknown, and he process by which *S* decides what value to settle for is certainly less clear than the corresponding method used by *E* in computing the threshold from a set of data obtained by the staircase-method.

In addition to its efficiency, the staircase-method is very convenient for tracing the temporal course of changes in a threshold when these changes are reasonably slow. For example, the method has been used to great advantage by Blough in plotting the temporal course of dark-adaptation in pigeons.[6]

There is a theoretical consideration which reveals a very important property of the staircase-method. An *S*'s response to a given stimulus in a series of stimuli depends not only upon the value of the given stimulus, but also to some degree upon the values of all of the stimuli that have already been presented, and upon the responses he has given to these stimuli. This fact results in phenomena variously called anchoring effects, series-effects, adaptation-level phenomena, perseveration, anticipation, etc.[7] As a consequence, the measured value of a threshold or a point of subjective equality (*PSE*) depends to some degree on the method of measurement employed.[8]

---

[6] D. S. Blough, Dark adaptation in the pigeon, *J. comp. physiol. Psychol.*, 49, 1956, 525-530; Method for tracing dark adaptation in the pigeon, *Science,* 121, 1955, 703-704.

[7] See Woodworth and Schlosberg, *op. cit.*, 225-233 for a general discussion of these effects.

The dependence of these measures (threshold and *PSE*) upon the method used must be carefully considered in the design of an experiment. For instance, consider an experiment to determine the effect of Condition X upon a *PSE*. Two *PSEs* will be determined, one with and one without Condition X. It is obvious that the same method, and in fact the identical measuring conditions (*e.g.* the same set of stimuli for the method of constant stimuli) must be used in determining each *PSE* if a discovered difference is to be attributed to Condition X. But it is evidently not so commonly acknowledged that the actual magnitude of the difference between the two *PSEs* will be influenced by the measuring method even when the two conditions are made identical. It is usually found that serial effects result in the *PSEs* being pulled toward the middle of the set of stimuli that are presented.[9] Therefore, if the two sets of stimuli delivered under the two conditions are the same, each of the *PSEs* will be pulled toward the *same* value by the serial effects. The measured difference between *PSEs* must therefore become smaller as the interactions between responses become greater.

Disadvantages. When the staircase-method is used, *S* is very much aware of the way in which the stimuli are being ordered. Even if he is naïve to begin with, it does not take him long to become aware of the procedure. If the judgments are easy this awareness is not disturbing, but as soon as the judgments become difficult, as they are in almost all psychophysical research, it becomes distressing.

When *S* has reported 'no' four successive times, his attitude and criteria on the next trial are not at all what they were at the beginning of the series. This is, of course, a reflection of the inter-dependencies of the series. Another related consequence of the series inter-dependencies is the fact that *S* can, if he pleases, manipulate the results at will. If he wished to 'cheat,' he could at any point in a series of stimuli and continue to give what look like meaningful data. The fact that it is possible for *S* willfully to malinger means that the results may also be influenced by the biases of an 'honest' *S*. This factor becomes crucial when effects of small magnitude are being studied, particularly when the *Ss* are acquainted with the purposes of the study. In many psychophysical studies, it is desirable to gather very large quantities of data on a few individuals, and, particularly when those individuals serve as *E* as well as *S*, the possible effects of *S*'s biases can be severe enough to offset the advantages of the staircase-method.

The double staircase-method. To reduce the interdependencies and to lessen the possibilities that they may bias the results, stimuli chosen by some procedure different from the procedure that determines the regular stimuli, may be inserted at various points in the series.

---

[8] Some of the forced-choice methods may be exceptions to this rule. (H. R. Blackwell, Contrast thresholds of the human eye, *J. opt. Soc. Amer.*, 1946, 36, 624-643). Heinemann has developed an ingenious way to use the staircase-method, in combination with a forced-choice procedure for the determination of differential thresholds (E. G. Heinemann, The relation of apparent brightness to the threshold for differences in luminance, *J. exp. Psychol.*, 61, 1961, 389-399).

[9] Woodworth and Schlosberg, *op. cit.*, 225-233.

One way of doing this is to scatter stimuli throughout the series with randomly selected values. Such a procedure would, however, have two disadvantages: (1) it would lose efficiency; and (2) the values of the non-staircased stimuli would so influence the final level that their selection would reintroduce artifacts of anchoring. A better solution is to run two series of staircased stimuli concurrently. The data resulting from a simple alternation of two staircases are shown in Fig. 3. The procedure is as follows. The *E* predetermines two starting points instead of the usual one. The first stimulus is presented at one of these predetermined levels and the response is recorded. On the next trial, the second predetermined level is presented, and the response is recorded. If the response to the first stimulus is 'yes,' the third stimulus level is made one step less intense than the first, and the response is recorded. The intensity of the fourth stimulus is determined by the response to the second stimulus, and so on. In this way, two staircase-series are run concurrently, one on odd- and the other on even-numbered trials, each alternate stimulus depending upon the response to the previous stimulus in its own staircase.

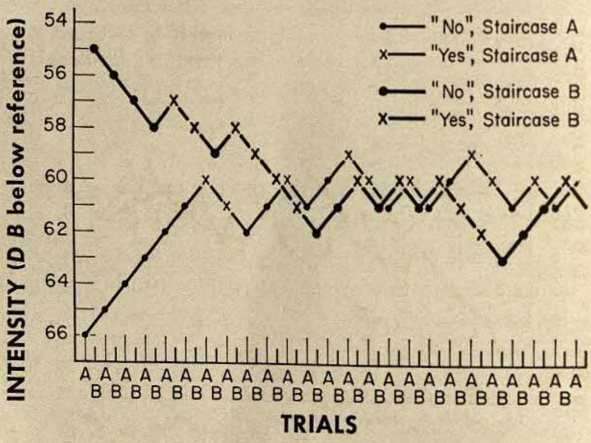


FIG. 3. DATA FROM A THRESHOLD-DETERMINATION IN WHICH TWO STAIRCASE-SERIES ARE ALTERNATED

Now go one step further. Let us call one staircase *A*, the other *B*. Instead of running two series, the *A* on odd and the *B* on even trials, the choice as to which staircase will be represented on any given trial is made randomly. Data collected from such a double randomly intermixed staircase are shown in Fig. 4. The order of occurrence of the staircases was chosen in advance from a table of random numbers. In general, if the two series are started apart as in this figure, they will come together and then cross and recross each other thereafter. The data to be used in the final computation of the threshold-value may be chosen in just the same way as they are for a single staircase. A double run may be considered either as two replications of the same condition, or the data may be combined and a single threshold-value computed.

When the random, double staircase-method is used, *S* feels none of the constraint that goes with the single staircase-method. The difference is perhaps best illustrated by the fact that it is no longer possible for *S* to close his eyes and give meaningful data. As long as the choice of staircase *A* or *B* is made randomly on each trial, there is no strategy of responding that *S* can follow which will prevent the stimulus-values from drifting gradually further and further up or down in a random walk. If, in fact, the two series do come together and then run along more or less horizontally, *S must* be responding to some aspect of the stimulus itself. There is no other way in which the stimulus-values can be kept from drifting. This means that the possibility of the effects of series- interdependencies biasing the computed threshold-value is greatly reduced.

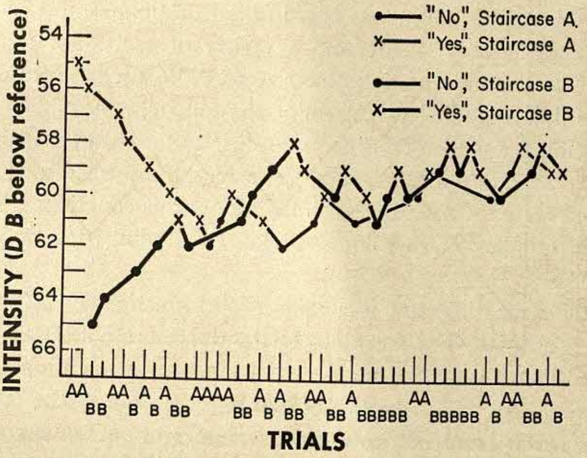


FIG. 4. DATA FROM A THRESHOLD-DETERMINATION IN WHICH TWO STAIRCASE-SERIES ARE MIXED RANDOMLY

In view of all the arguments presented above, the double-staircase method provides a means of determining psychophysical functions that has the practical advantage of high efficiency while retaining the theorectical advantage that the final results are relatively little influenced by the biases that *E* must introduce into any experiment.

University of California, Berkeley

TOM N. CORNSWEET

USES OF THE SEMANTIC COUNT IN EXPERIMENTAL STUDIES OF VERBAL BEHAVIOR

In many studies of verbal behavior, it would be helpful to know how the usage of a given word is apportioned among its various meanings. We wish here to call attention to a source that provides such information but is not widely known among those who might wish to use it. In addition to describing this semantic count, we shall suggest a few purposes for which it may be employed, and shall show one use made of it.

A semantic count was made in the 1930s under the direction of Lorge and Thorndike.[1] This count was based on an original sample of over two million running words taken from a variety of texts. For some of the words, a second sample of the same size was also used. Every occurrence of each word was studied to determine the apparent meaning in context, and the meaning was then classified according to the list of meanings given in *The Oxford English Dictionary*.[2] The most frequently used words in the language were, in most cases, omitted from this count. In 1949, Lorge published a supplementary volume, giving the findings for the 570 commonest words.[3]

While the semantic count is a mine of information, extracting the information is arduous. The meanings of the words are not stated but rather are encoded in terms of the numbers and letters employed in *The Oxford English Dictionary*. To make use of the semantic count, the reader must therefore have at hand not only the 6 volumes of *A Semantic Count of English Words* and the supplementary volume, *The Semantic Count of the 570 Commonest English Words,* but also the 13 volumes of the 1933 edition of *The Oxford English Dictionary*. Furthermore, the distinctions among meanings are extremely fine; thus, 43 meanings are given for the word 'man.'

To encourage wider and easier use of the data of the semantic count, West prepared what is essentially a simplified abridgment, *A General Service List of English Words*.[4] This volume contains about two thousand main entries, to which some derived forms are added. The meanings in the semantic count of Lorge and Thorndike have been regrouped more coarsely into main areas of meaning. For example, only eight meanings are listed for the word 'man.' Each meaning is stated briefly or indicated by an example; sometimes both meanings and examples are given. The relative frequency of each meaning of a word is given as a percentage of the total usage of the word. Here, for example, are the meanings given for the word 'man' and their respective percentages of employment:

---

[1] Irving Lorge and E. L. Thorndike, *A Semantic Count of English Words,* 1-6, 1938.

[2] *The Oxford English Dictionary,* 1-13, 1933.

[3] Lorge, *The Semantic Count of the 570 Commonest English Words,* 1949.

[4] Michael West, *A General Service List of English Words,* 1953.

human, often with a capital letter	10 %
adult	14
male	9
soldier	6
servant	1
vaguely, person	54
verb	0.4

As in the case of most entries, the percentages do not add up to 100; certain minor meanings having been omitted.

Now let us note a few ways in which this information may be used in studies of verbal behavior. We will give examples with special relevance to studies of association of words but with application to verbal learning as well. (It should be remarked here that, as we have shown in our previous Note,[5] there are errors in the total frequencies given for some words by West. There may, then, also be errors in the data on relative frequencies of meanings. This may lead to some inaccuracies in any study employing this material, but it is unlikely to endanger conclusions based upon the use of a substantial number of items.)

(1) Selection: Choice of stimuli according to part-of-speech. This poses problems in English, because many common English words are employed as more than one part of speech. For example, 'command,' 'hand,' 'sleep,' 'whistle,' and 'wish' may each be used either as a noun or as a verb. Kent and Rosanoff prepared their well-known list of stimuli to include only nouns and adjectives.[6] The five words mentioned above were included as nouns. Should they have been? According to the semantic count, 'sleep,' 'whistle,' and 'wish' are employed more often as verbs than as nouns; this makes questionable their inclusion in the list.

(2) Part of speech: Determining the part of speech of isolated verbal responses. In certain studies of association, for example, it is necessary to determine whether each response is of the same part of speech as the stimulus. Judgments of part of speech have usually been made intuitively by the experimenters, but as we have seen in *(1)* above, such judgments may

---

[5] M. R. Rosenzweig and David McNeill, Inaccuracies in the semantic count of Lorge and Thorndike, this JOURNAL, 75, 1962, 316-318.

[6] G. H. Kent and A. J. Rosanoff, A study of association in insanity, *Amer. J. Insanity,* 67, 1910, 37-96, 317-390.

be in error. The semantic count provides a criterion for decision, or at least a measure of the error that may be involved in the classification of a particular word.

(3) Meaning: Predicting the meaning of a word presented in isolation. Most common words have several different meanings. Is it posible to predict in what sense a person will take such a word when it is presented in isolation or with reduced context as, for example, in a list of unrelated words to be learned or in a test of association? We wish to propose here that the sense in which the word will be taken when presented in isolation is that meaning in which the word is employed most frequently in the language (the *predominant* meaning, as we shall call it).

To test this hypothesis, let us examine data from a test of word-association. The most frequently given associative response to a word (called the *primary* response) provides an indication of the sense in which the word is taken in isolation. For example, the primary response to the stimulus 'man' is 'woman.' Among the meanings of 'man' listed above, this response corresponds with 'male,' which is not the predominant meaning. Let us determine in what proportion of cases the primary associative response to a word corresponds with the scope of the predominant meaning of the word. The widely used list of stimulus-words of Kent and Rosanoff consists of 100 rather common words, 87 of which appear in West's book. Twenty-six of these words are listed with only a single meaning, so no doubt about the meaning arises in these cases. We are left, then, with 61 cases to test our hypothesis. The primary responses will be taken from the most extensive recent set of norms for response to these stimuli, the norms of Russell and Jenkins.[7]

TABLE I

DISTRIBUTION OF AGREEMENTS BETWEEN PRIMARY RESPONSES AND PREDOMINANT MEANINGS ACCORDING TO THE PERCENTAGE OF USAGE OF THE PREDOMINANT MEANING

	% usage of predominant meaning			
	≥60	40–59	20–39	Total
Number cases	25	19	10	54
% Cases in which predominant meaning and primary response agree	100	63	50	78

In 54 cases of the 61, the primary response could be assigned with little doubt to one of the meanings of the stimulus-words. In the other seven cases, the response corresponded to more than one meaning of the words, so no decision for or against the hypothesis could be reached for these items. (Among these ambiguous cases are five names of colors. Each of these occurs over 60% of the time as an adjective and the rest of the time as a noun. The response, in each case another color-word, therefore agrees with both the adjectival and noun meanings and so cannot be used to test the hypothesis.)

---

[7] W. A. Russell and J. J. Jenkins, *The Complete Minnesota Norms for Responses to 100 Words from the Kent-Rosanoff Word Association Test*, 1954.

The primary response was found to be related to the predominant meaning in 42 of the 54 critical cases. Thus, when *Ss* saw the words in isolation in a word-association test, their most common responses corresponded with the area of the predominant meanings in 78% of the cases.

In Table I items are classified according to the percentage of usage accounted for by the predominant meaning. A further fact emerges from this analysis: The greater the proportion of usage devoted to the predominant meaning, the greater the probability that the primary response will agree with the predominant meaning. For the items examined, when the predominant meaning accounts for at least 60% of the usage of a word, the primary associative response is related to the predominant meaning in every case. As the percentage of usage falls off, so does the percentage of agreement, but it should be observed that the percentage of agreement exceeds the percentage of usage for each of the three ranges.

In demonstrating that a word presented in isolation is usually taken in the sense of its predominant meaning, we believe that we have given a further indication of the value of the semantic count to students of verbal behavior.

University of California, Berkeley

MARK R. ROSENZWEIG
DAVID MCNEILL

SOME METHODOLOGICAL PROBLEMS IN MICROGENETIC INVESTIGATION

In recent years attention has been directed to a consideration of microgenesis[1], which Flavell and Draguns, who recently reviewed the work done in this field, defined as follows:

> The term "microgenesis," first coined by Werner as an approximate translation of the German word *Aktualgenese,* will refer here to the sequence of events which are assumed to occur in the temporal period between the presentation of the stimulus and the formation of a single, relatively stabilized cognitive response (percept or thought) to this stimulus. More specifically, the term will refer primarily to the prestages of extremely *brief* cognitive acts, *e.g.* the processes involved in immediately perceiving a simple visual or auditory stimulus, conceptually generating a word association, etc.[2]

In summarizing the results of the European studies, they describe the microgenetic sequence as follows:

> When a perceptual stimulus is presented under conditions of gradually increasing clarity, the initial perception is that of a diffuse, undifferentiated whole. In the next stage figure and ground achieve some measure of differentiation, although the inner contents of the stimulus remain vague and amorphous. Then comes a phase in which

---

[1] Heinz Werner, Microgenesis and aphasia, *J. abn. soc. Psychol.,* 52, 1956, 347-353.

[2] J. H. Flavell and Juris Draguns, A microgenetic approach to perception and thought, *Psychol. Bull.,* 54, 1957, 197-217.

contour and inner content achieve some distinctness and a tentative, labile configuration results. Finally, the process of Gestalt formation becomes complete with the addition of elaborations and modifications of the 'skeletal Gestalt' (*Gestaltgerüst*) achieved in the previous stage.[3]

There have been many attempts to integrate microgenesis into general developmental theory. The data obtained by Philips and Framo involving tachistoscopic exposure of Rorschach cards, support such a view.[4] The sequence of events of their normal *Ss* closely approximated the findings of Hemmendinger in his developmental study (children 3 to 10 yr. old) of the perceptual non-projective aspects of response to the Rorschach cards.[5]

Werner and Kragh have also noted the similarities between the microgenetic and developmental sequences and have argued in fact that the microgenetic sequence recapitulates development.[6] Kragh's argument is based upon his study involving *TAT* cards, but his evidence does not necessarily follow from an examination of the protocols reproduced in his book, at least not according to the authors of this paper. Whereas Werner is on reasonably safe ground concerning himself with the structure of the responses as being indicative of a given developmental level, Kragh's notion of recapitulation is largely historical, necessitating careful checks on the past history of the *Ss*. Kragh, therefore, really addresses himself to an individual clinical problem rather than to general considerations of development and developmental theory.

If the above findings and their interpretation are accurate, then an area of investigation has been opened which may help to solve some of the mysteries of the development of percept-structure. An examination of the methodology employed in microgenetic experiments, however, shows that any results obtained may be called into question on three counts: (1) confounding of the main independent variable; (2) biasing of stimulus-structure, and (3) possible bias due to the type of response elicited from the *Ss*.

Investigations of microgenesis have usually involved presenting stimuli for very brief time-intervals or at low illumination, and progressively increasing the time-intervals or illumination in discrete steps. The assumption underlying this method is that it succeeds in dissecting the ordinary course of events in the real-life situation, so that it is possible to observe the characteristics of the early and intermediate stages in the development of a single percept. All of the studies, however, take *S* through the full range of exposures, thus introducing at each exposure a repetition of all preceding stages. Any results obtained and, of course, inferences based upon them, cannot, therefore, be attributed to systematic variation in the stimulus-dimension itself but must be the result of the confounding of such changes with whatever effects there may be of repeated presentation. The only way of avoiding such confounding, and therefore more faithfully to reproduce the structure if not the content of the real-life situation, would be to obtain independent responses at each time-interval or intensity of illumination. This procedure requires either that the same *S*s respond to different stimulus-objects at every presentation or that different *S*s respond to a single stimulus-object at every presentation. The former case still leaves the problem of the effects of practice in perceiving under reduced conditions while the latter eliminates repetition effects completely. It is true that such methods make impossible an accurate record of the development of a single percept in an individual, but, it is believed, experimentation in microgenesis is logically incapable of producing such a record, solely as a function of the independent variable being manipulated. Even if it were possible, by some means, to stretch the period of micro-development from 150 m.sec. to 150 sec., the very fact that *S* is called upon to respond at successive points in the time-continuum adds a confounding set of stimulus-variables feeding back from the responses made—a state of affairs which does not appear to occur in the ordinary life situation. It would seem, then, that investigators would have to be satisfied with structural data for the most part and composite content from cross-sectional groups of *S*s if they are to study the effects of their experimentally induced independent variable in isolation.

---

[3] *Ibid.*, 198.

[4] Leslie Phillips, and J. L. Framo, Developmental theory applied to normal and psychopathological perception, *J. Pers.*, 22, 1954, 465-474.

[5] Lawrence Hemmendinger, Perceptual organization and development as reflected in the structure of Rorschach test responses, *J. proj. Tech.*, 17, 1953, 162-170.

[6] Heinz, Warner. The concept of development from a comparative and organismic point of view, in D. B. Harris, (ed.) *The Concept of Development*, 1957, 125-148; Ulf Kragh, *The Actual-Genetic Model of Perception-Personality*, 1955.

A second methodological problem involves the physical properties of the stimulus-figures used. In the studies reported by Flavell and Draguns the figures used have been, for the most part, those with continuous outlines.[7] The general finding of global or whole figure-ground responses reported by these investigators could well be a function of the limited class of figures used. Figures with continuous contours represent one of many classes of figures and it is not inconceivable that figures constituted of discontinuous outlines or of discrete elements would result in different responses. If the Gestalt oriented investigators of perceptual development are correct concerning the organized nature of perception, the construction of the stimulus-pattern should make no difference to the sequence of events. At least if the stimulus-figures were composed of discrete elements there would be the possibility of an empirical test of the implications of their position.

---

[7] Flavell and Draguns, *op. cit.*, 197 f.

The third methodogical problem involves the structure and the content of the behavioral response required of *S*. Flavell and Draguns report that all but one of the studies which they reviewed employed some form of verbal description.[8] The exception involved an attempt to reconstruct the original stimulus-pattern. The work of linguists such as Whorf suggests that the experimentalist must take note of the classifying and organizing properties of his *S*s' linguistic code.[9] It has been argued that the ready made classifications of the language in use determine (in part) as well as direct, perceptual, responses. Our own linguistic code would tend, therefore, to bias responses in favour of organized wholes by virtue of forcing the *S*s to identify the vague stimulus-pattern as a member of some class of objects. It seems desirable, therefore, to choose some form of response which avoids such biasing. A possible form of behavior is line-drawing reproduction of the stimulus-pattern by *S*. Although such behavior effectively avoids linguistic factors, drawing involving continuous lines might also bias responses towards wholeness. In making use of the stimulus-patterns constructed of discrete elements it would be desirable that *S* should be requested to reproduce the pattern employing materials essentially similar to those with which the figure was constructed. This procedure would have the added advantage of making the content and structure of the response similar to that of the physical pattern. It would also avoid any extraneous influence toward 'whole' responses.

A pilot study of the microgenesis of form-perception was conducted recently by the authors. Because no tachistoscope sensitive enough was available, the major independent variable selected was increasing brightness. The study employed 246 naïve *S*s. The principles discussed above were adhered to in that independent *S*s were used at every level of brightness level. The stimulus-objects were transparencies of figures constructed of marbles and the response elicited from every *S* was a reproduction of the stimulus-pattern constructed of marbles placed on a marble board. Although this investigation was meant to serve only as a preliminary means of discovering relevant dimensions and for testing the method, the results indicated overwhelmingly that the first stage in the microgenetic sequence is not an undifferentiated whole response but consists of a reproduction of intersecting figural components, *e.g.* angles and crosses. A global whole appeared at a later stage. Kragh in the tabular report of his 'line-akt' series, also noted that the large majority of his *Ss*' initial responses were of this same kind. "A detail or details of the correct object (*e.g.* 'two lines at angle') are marked by *f* in the table; *S* should then have perceived only a 'geometrical' *fragment* of the figure."[10] Thirty-five out of 51 *Ss* were scored *f* on their first response to at least one of the stimuli presented. It must be pointed out that Kragh obtained such results from verbal responses. Perhaps these results would have approximated the unanimity of the present authors' findings (48 out of 50 component responses) had he used a reproductive rather than a verbal response. In any case there appears to be some doubt as to whether Flavell and Draguns are correct in their description of the microgenetic sequence.[11]

---

[8] Flavell and Draguns, *op. cit.,* 197 f.

[9] B. L. Whorf, *Language, Thought, and Reality,* 1956.

University of Sydney
Australia

R. H. POLLACK
MARY ROSE CHAPLIN

THE EXPERIMENTAL UPSURGE, 1797-1874

Boring remarks, in effect, that the experimental movement becomes important from about 1860 on. A usual date for its public recognition is the opening of the Leipzig laboratory in 1879.[1] In the course of a wider study of the preëxperimental period, taken as 1797-1874, it has seemed worth while to compare the progress of experimental with that of some other kinds of psychology. While the results cannot claim to be definitive, they seem, at this point, to justify an interim note.

Production from 1797 to 1874 was assessed by extracting the relevant entries in Rand's *Bibliography*.[2] An entry, for this purpose, is any work identifiable by date with reasonable ease (including short-run periodicals). Second or subsequent editions, reprints, translations, critical articles, and reviews, unless quoted in their own right, are excluded. This note will attempt some comparison of the section on "Experimental and Physiological" with the sections larger in size and nearest in kind.

The section "Experimental and Physiological" has 140 entries from 1797 to 1874 inclusive. It is the third largest section, exceeded only by a kindred section, "Sight" (153 entries), and the somewhat amorphous one, "Systems and Essays" (389). Next in descending order come "Comparative" (128); "Mind and Body" (96); "Individual and Social" (89); "Abnormal and Clinical" (88); and "Insanity and Idiocy" (82). These are the larger sections, as the next, "Nervous System," 68, shows a big drop in entries. Nearest in kind to "Experimental," from the titles, besides "Sight," are "Nervous System,"; "Perception" (11 entries); "Psychophysics" (7); and "Work and Fatigue" (4).

---

[10] Kragh, *op. cit.*, 155.
[11] Flavell and Draguns, *op. cit.*, 197 f.

[1] E. G. Boring, *A History of Experimental Psychology*, 2nd ed., 1950, ix, 323.
[2] Benjamin Rand, Bibliography of philosophy, psychology, and cognate subjects, in J. M. Baldwin (ed.), *Dictionary of Philosophy and Psychology*, 3, 1905, 917-1192.

The entries in the various fields are distributed in time by decades beginning with 1797-1806, as follows:

		'97–'06	'07–'16	'17–'26	'27–36'	'37–'46	'47–'56	'57–'66	'67–'74
Experimental and Physiological:	No.	2	3	4	9	18	24	31	49
	%	1.4	2.1	2.9	6.4	12.9	17.1	22.1	35.0
Sight:	No.	0	7	4	8	7	27	40	60
	%	0	4.6	2.6	5.2	4.6	17.0	26.1	39.2
Systems and Essays:	No.	11	9	27	46	48	74	76	98
	%	2.8	2.3	6.9	11.8	12.3	19.0	19.5	25.2
Comparative:	No.	3	3	6	7	10	11	34	54
	%	2.3	2.3	4.7	5.5	7.8	8.6	26.6	42.2
Mind and Body:	No.	1	2	2	3	13	22	27	26
	%	1.0	2.1	2.1	3.1	13.5	22.9	28.1	27.1
Individual and Social:	No.	3	1	0	7	5	17	22	34
	%	3.4	1.1	0	7.9	5.6	19.7	2417	38.4
Abnormal and Clinical:	No.	2	2	13	9	17	10	21	14
	%	2.3	2.3	14.8	10.2	19.3	11.4	23.9	15.9
Insanity and Idiocy:	No.	6	4	6	7	14	11	17	17
	%	7.3	4.9	7.3	8.5	17.1	13.4	20.7	20.7

These large sections show that "Experimental" has a considerable concentration of production in the last two decades, with a quite steady, rather stepwise rise toward this higher level. "Experimental" is of course intertwined with "Physiological," and a breakdown of this section, in progress, may show that purely experimental development was more erratic. Certainly this is so with the kindred section, "Sight," which shows higher late-decade concentration, and a more fluctuating early course. "Systems and Essays," as befits its *omnium gatherum* character, shows pretty steady increase in production with earlier concentration. In "Mind and Body," development is by steps with rather less concentration at the end and with a different pattern (slight tapering off). Not surprisingly, "Abnormal" and "Insanity" belong to a different 'family' of sections—those devoted to standing problems, rather than developing new ones. The last two decades show less concentration, though "Insanity" progresses more stepwise than "Abnormal," which fluctuates throughout. "Individual and Social" is of another kind again with high concentration in the late decades, and downward trends over the first three and the fourth-fifth decades.

Now for the other kindred groups:

Nervous system:	No.	1	2	1	3	12	7	22	20
	%	1.5	2.9	1.5	4.4	17.6	10.3	32.4	29.4
Perception:	No.	0	0	0	0	1	1	1	4
Psychophysics:	No.	0	0	0	0	1	0	4	2
Work anf Fatigue:	No.	0	0	0	0	0	0	0	4

The last three are too small for percentages to be helpful or significant. "Work and Fatigue" is very new; "Psychophysics" scores only during the Fechner decade, 1857-1866; "Perception" began to develop in the fifth of our decades, 1837-1846 (the actual date is 1846), and "Nervous System" is a step-development with concentration in the later decades.

From these findings, we draw the following conclusions: (1) Boring's notion of an experimental upsurge at about 1860 is justified. The actual entries for 1860-1874 total 75, or 53.6%, though a few earlier single years (*e.g.,* 1852 and 1855 with 6 entries each) exceed any other year until 1866. (2) Late concentration, with steady stepwise progress, characterize "Experimental," "Sight," "Mind and Body," and "Nervous System," in varying degrees. We can hardly compare the late-starting small sections. "Systems and Essays" follows the same broad pattern, but what this signifies remains to be seen. (3) "Abnormal" and "Insanity" are in a different family, for these problems are ever with us. (4) "Individual and Social" is on its own too, though less markedly. Note that the choice of 1797-1874 as the first decade means that the last decade is only 8 yr. and that concentration is, therefore, slightly understated. For reasons belonging to the more general inquiry, it is best, however, to stick to 1797-1874.

This note, then, places "Experimental and Physiological" in relation to the other large and kindred sections. "Experimental" is a member of the moderate *nouveau-riche* family in psychology, by contrast with the old residents in the psychological household, like "Abnormal" and "Insanity." Not all the family history is revealed here, but further inquiries are proceeding.

University of Tasmania
Hobart, Tasmania

J. A. Cardno

THE APPARENT DISTANCE OF MOUNTAINS

Of the various illusions connected with distance-perception, perhaps the most often observed is the fact that a mountain which is in fact very far off may appear to be quite near. A common story is that of the unwary vacationer who starts an afternoon walk to a summit 50 miles away. The usual explanation of the illusion is the clearness of the air in mountainous country, that is, the absence of aerial perspective which would ordinarily provide a clue to its distance. The theory is that, for a patch in the visual field, its degree of blueness, of loss of sharp contour, or loss of detail, or haziness (descriptions vary) is an indicator of the distance of the corresponding object in the environment, and also that this sign of distance has been learned and is relied upon by the normal observer. The explanation and the theory have been so often stated that whenever the illusion is observed it is taken to confirm them. The degree of validity of this clue, however, is not known, and the present writers have some doubt about it.[1]

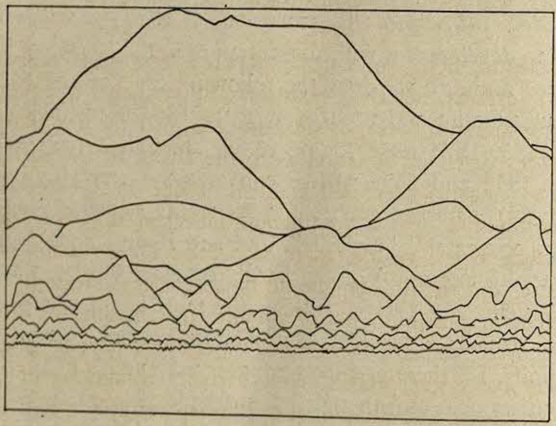


FIG. 1. THE ILLUSION OF APPARENT DISTANCE

The purpose of this note is to suggest another possible explanation of why a distant summit can look nearer than it is. In level country, increasing terrestrial distance is correlated with the decreasing angular size of elements upward in the optical array (or in a pictorial projection of the array) since the terrain features tend to have the same size. But in the neighborhood of a mountain, the distant earth-shapes may be much larger than the nearer ones and the usual optical gradient will then be altered. To dramatize this change in the gradient, the accompanying illustration (Fig. 1) shows an extreme instance, where supposedly the actual sizes of the earth-shapes strongly increase with distance. There results an increasing size of the corresponding optical elements upward in the array instead of the decrease that would occur with the usual terrain. The normal gradient has been reversed. The clue of 'perspective-size' in the picture is that for a ceiling, not that for a level ground. The distant mountains in the picture do in fact seem strangely near.

---

[1] J. J. Gibson, *The Perception of the Visual World*, 1950, 114-116 and 141.

For the continuous surface of the terrain immediately surrounding an obesrver, concurrent gradients of disparity and of motion parallax will normally accompany the perspective of optical texture and the former will provide true information. For a distant mountain with its foothills, however, the parallax between the two ocular station-points and that between successive positions of the head will often be too slight for detection, and the gradients will be indistinguishable from those corresponding to a flat picture.

The illustration, being an outline drawing, represents only the edges of terrestrial surfaces, not their textures. The gradients of optical texture corresponding to forest cover, or sometimes to the structure of rock slopes, would provide partial information as to the layout of the surfaces which would tend to cancel the reversed gradient of size. In the special case of a mountain whose face was a continuous unbroken slope of forest or of rock, the whole being visible, there would be a continuous gradient of optical texture, and the illusion should be absent or decreased.

There are many possible types of stimulus-information for distance in a natural optic array, some of them based on constant laws of geometry and optics, others based on probable but variable conditions of geological features, atmosphere, and illumination. Illusory perception will depend on the combination of circumstances which holds for the particular situation, and also probably on the degree to which the attention of the observer has been trained to register the reliable information in the array. The popular idea that the clearness of mountain air explains the misperceptions of distance is too simple.

Cornell University

JAMES J. GIBSON
HOWARD FLOCK

CONTEXTUAL INFLUENCE UPON THE PERCEPTION OF COLOR

Students of the vagaries of vision may be interested in an observation made by the writer quite by chance. The only materials necessary are a sheet of blue-ruled yellow paper from a legal pad and a sheet of plain white paper. Also necessary is a white light source, such as is provided by most fluorescent lamps.

The procedure to be followed consists simply in laying the white sheet upon the yellow one with one edge of the former parallel and close to the second blue line of the latter. That line then appears to be green while the top line remains blue; furthermore, the yellow portion of the lined sheet adjacent to the edge of the white sheet takes on a deeper, more saturated hue. The effect is somewhat like that obtained by laying a sheet of yellow cellophane upon the lined sheet. If the white sheet is moved slowly down toward the bottom of the yellow sheet, the effect is heightened.

College of Medicine
University of Illinois
Chicago, Illinois

CHARLES LIGHTFOOT

THIRTY-THIRD ANNUAL MEETING OF THE EASTERN PSYCHOLOGICAL ASSOCIATION

The Eastern Psychological Association met April 26-28, 1962 at the Chalfonte-Haddon Hall in Atlantic City, New Jersey. The registration at the meetings totaled 2008. Of this number, 991 were regular members of the Association and 1017 were either guests or newly elected members. The present membership of the Association currently totals 3332.

A committee headed by Arthur I. Siegel was responsible for local arrangements. The program was planned by a committee under the chairmanship of Henry P. David. A total of 291 papers was presented in 50 sessions. Also scheduled were 6 symposia, 4 special sessions, 3 special group-meetings, and 2 films.

The majority of paper-reading sessions was devoted to topics in experimental, comparative, and physiological psychology. They were more specifically the following: operant behavior, animal learning, discrimination learning, conditioning, perceptual learning, verbal learning, human learning, efficiency and verbal reinforcement, cognitive processes, decision processes, verbal behavior, perception, after-effects and imagery, visual perception, vision, audition, sensory processes, animal motivation, avoidance behavior, animal behavior, psychopharmacology, physiological, and brain mechanisms. Four sessions were devoted to developmental and child behavior. Social psychology was covered by sessions entitled social behavior, social attitudes, and social perception. Papers in clinical psychology were grouped under schizophrenia, clinical evaluation, clinical judgment, projective techniques, and psychotherapy and psychosomatics. Other individual papers related to personality, personality and cognition, evaluation and measurement, industrial and personnel, and engineering psychology.

The symposia were entitled "Instinct, early experience, and the behavior disorders," "Psychometric implications for the localization of brain function," "Reinterpretations of the Schreber case: Freud's theory of paranoia," "Inter-nation perceptions, communications and reduction of conflicts," "Changes in organizational structure: parameters, attitudes, and processes," and "Research problems in schizophrenia."

Special sessions were devoted to discussion of the report of the Joint Commission on Mental Illness and Health, the processes of cognitive growth, therapy with children, and a new projective test (the Hand Test). Psi Chi invited Josef Brozek to speak on Soviet experimental studies of diet and behavior. Special group meetings were held by Psychologists Interested in the Advancement of Psychotherapy and by the Society for Client-Centered Counseling.

George A. Miller presented the annual presidential address, "Some psychological studies of language."

During the business meeting, Richard L. Solomon was announced as the President-elect and David C. McClelland and Eliot Stellar as new members of the Board of Directors. Martin M. Bruce was elected to a second term in the office of Treasurer. Also reported, on the basis of extensive revisions of the by-laws (approved earlier by mailed ballot), were the incorporation of the Association and a change in the office of Secretary to that of Executive Secretary. The membership endorsed a decision of the Board of Directors to donate $2000 of accumulated surplus funds to a fund for aiding young foreign psychologists to attend the International Congress of Psychology in August, 1963, in Washington, D.C.

The 1963 meeting of the Association will be held at the Statler-Hilton Hotel in New York City, April 11-13. Symposia will be scheduled for evening hours and research papers for daytime sessions, the program to begin at 2 P.M. on Thursday.

Garden City, New York MARVIN A. IVERSON

THIRTY-SECOND ANNUAL MEETING OF THE ROCKY MOUNTAIN PSCHOLOGICAL ASSOCIATION

The thirty-second annual meeting of the Rocky Mountain Psychological Association was held at the Finlen Hotel in Butte, Montana, on May 10–12, 1962. Approximately 100 persons attended.

The program included both papers and symposia. A highlight of the meeting was the invited address by Dr. T. C. Yu of the School of Journalism, Montana State University, on "Some observations on psychology and international situations." The presidential address by Dr. Bert Sappenfield was entitled "The tyranny of terminology."

Officers for the coming year are: President, Earl Swartzlander; President-elect, Merrell Thompson; Secretary, Wilbur C. Miller; Treasurer, Richard Lake.

The 1963 meeting will be held in Cody, Wyoming, with Wilson Walthall, Jr., University of Wyoming, serving as convention manager and W. Harry Sharp serving as program chairman.

University of Denver WILBUR C. MILLER

Carney Landis: 1897-1962

The death of Carney Landis on March 5, 1962, ended a distinguished and exceptionally productive research career. Landis' laboratory was his love and he was unusually faithful to it. Not a man of great physical energy or stamina and frequently plagued by nagging ill health, he set himself a driving pace which often placed him at his desk in the early morning hours that he might have some undisturbed time for study and writing before the routine day's work began. Even during the painful and debilitating illness that eventually ended his life he kept steadily at work when he could. His work was his life.

Landis was born on January 11, 1897, in the small Ohio town of West Alexandria. His boyhood was typical of the place and time and included the inculcation of a simple but rigid Christian ethic, ill adapted to the sophisticated complexities of the urban and professional atmosphere in which he later found himself. Much of the non-didactic part of his subsequent psychoanalysis was concerned with the working through of these problems. For him, however, no ethical problems existed in the laboratory. Here the situations were well defined, the rules explicit, and human acceptance practically unanimous. The results always could be predicted —ultimately good won out and evil was vanquished. Whatever the complexities and ambiguities of one's responsibilities to society, one's responsibilities to data were simple and unequivocal. In the laboratory Carney Landis could live at peace with his early religious upbringing and his subsequent experimental training.

The facts of his professional life are deceptively simple. After a brief period at Carnegie Institute of Technology he entered Ohio State University where he obtained his A.B. in 1921. The following year he earned his M.A. from Dartmouth while teaching there. Then came Minnesota where he acquired the Ph.D. degree under Karl Lashley in 1924. After two years on a National Research Council fellowship he was appointed in 1926 Assistant Professor at Wesleyan University, and there he was promoted to Associate Professor in 1928. It was during this period that he worked with the Behavior Research Fund in Chicago. In 1930 he accepted an appointment as research associate psychologist at the New York State Psychiatric Institute and Hospital in New York City. Here he remained until his death, becoming Chief Research Psychologist in 1958. He also had an appointment as Professor in the Columbia Department of Psychology, a part which satisfied his academic needs. His life included brief teaching visits to Harvard, UCLA, and the University of Washington, as well as a service mission to Colombia, where he lectured widely and was given an honorary Doctor of Science from the University of Antioquita in Bogota; but he had no real desire to be a college professor and preferred the research freedom of his laboratory.

By contrast his scholarly contribution was quite complex owing to his widely ranging interests. His combination of ceaseless drive and insatiable scientific curiosity involved him in research that ranged from physiology to phenomenology, and from the developmental study of the individual neurotic to the interpretation of mass statistics on mental disorder. His work was primarily empirical, for he had little aptitude and even less sympathy for theoretical matters. Coupled with this devotion to objective data went his lifelong enthusiasm for elaborate instrumentation.

During his undergraduate days at Ohio State, the naïve, dogmatic behaviorism of Weiss had given him the needed authoritative framework for his research. Later, Watson and Lashley—Lashley who remained a father figure to him throughout his life—added some details. Landis, however, was never a behaviorist in any narrow sense. The denying of consciousness or the arbitrary exclusion of any facet of human behavior seemed foolish to him. The existence of these facts was obvious to him. His task was to bring their study under objective conditions. Later on he would be sympathetic to operationalism and somewhat amused by logical positivism (he regarded the logic of science as largely an elaboration of the obvious). Let other men stop to rationalize; he was busily at work. In this attitude he was the fore-runner of an increasingly numerous group, the professional research psychologists, who accept their trade, concentrate their creative energies on the development of their investigative techniques, and spend their lives producing empirical answers to definable problems. Their existence is witness to the increasing professional maturity of psychology as a science and of its liberation (be it good or bad) from philosophy.

As a result of his fervent laboratory empiricism and his perpetual skepticism of accepted dogma, his early impact upon experimental psychology was extensive. Those were the booming twenties when all the ultimate answers lay just around the corner, the period of *The Psychologies of 1925,* of mass action and equipotentiality, when a good concept in the bush was worth half a dozen facts in the hand (why replicate the obvious?). In this atmosphere Landis' flat, empirically-oriented skepticism was an intrusion, like a snowstorm in a tropical paradise. Within his chosen area of abnormal psychology (defined then much more broadly than it would be now) conceptual edifice after conceptual edifice toppled before a series of well planned experimental attacks. His work on the expressive patterns of emotion cast doubt on their inherent nature. The galvanic skin reflex was being fought over by the various proponents of emotion, of intellectuation, of orexis, and of conation. Landis' answer was to put together in the same experiment stimulus-situations of all these sorts and to discover that the *GSR* occurred in *all* of them. His work with Page in *Modern Society and Mental Disease* helped to introduce a new and sophisticated approach in the interpretation of the statistics of mental disorder. His studies on *Sex in Development* reaffirmed the importance of the disease process in psychosis and challenged the supremacy of dynamic, developmental concepts.

As would be expected, Landis' relations with psychiatry were ambivalent. He could be caustic, as when he wrote of the average psychiatrist, "He has no adequate knowledge of scientific reporting, no comparable theoretical vocabulary, and but the haziest ideas of scientific logic." Yet he was fair when protesting the relative neglect of Freudian doctrine in the texts on abnormal psychology in the thirties. His own analysis, while not entirely didactic, was in large part undertaken in a spirit of scientific inquiry. He felt that to be fair in his criticism of psychoanalysis he should go through the experience himself. Often considered hostile to psychotherapy, he saw himself merely as being objective in stressing the lack of any adequate validating evidence of its efficacy.

Landis was not by nature an iconoclast. He was disturbed when he acquired a reputation for personal negativism. As his generation of psychologists grew older and skepticism became the rule, he felt more at home. The investigation of the startle-pattern particularly pleased him, for here was a simple, predictable pattern in which he could believe, and ultra-rapid photography with its magnification of time involved just the kind of complicated gadgetry that delighted him. Still his interests continued to range widely. His work on personality inventories laid the foundation for much of the subsequent work on selection instruments in World War II. His later work on psychosurgery, on flicker fusion, and on pharmacology was highly regarded and bore the typical Landis trademarks—extensive bibliographical preparation and elaborate instrumentation. In the last years of his life he was completing a study of the personal experiences of mental patients. It will be published posthumously as *The Varieties of Abnormal Experience*.

The influence of Landis' work probably is likely to have been more widely felt among his generation than will be noted in history. His firm insistence on the highest standards of scientific inquiry was a needed corrective influence for the times. Perhaps psychology needs a Landis less now, nor is it likely soon again to have an experimental psychopathologist of his special kind of excellence and ability.

Northwestern University WILLIAM A. HUNT

BOOK REVIEWS

Edited by T. A. Ryan, Cornell University

Handbook of Abnormal Psychology: An Experimental Approach. Edited by H. J. Eysenck. New York, Basic Books, Inc., 1961. Pp. xvi, 816. $18.00.

Fifteen of the 20 chapters of this book were written by members of the Department of Psychology, University of London. The other authors include three members of the Department of Psychiatry and a psychologist and physician on the staff of Maudsley Hospital. According to the editor the book is the fruit of many discussions among the authors, whose major concern was selection of the valid scientific material available in the field of abnormal psychology. The editor also states that the authors were highly selective, finding only about one fifth of the published material to be of satisfactory value. The purpose of the book is stated as that of bringing "the experimental tradition of psychology to bear upon the 'abnormal' behavior of the organism." The traditional approach to the material in terms of diagnostic categories is disavowed, and instead emphasis is placed on the defective functioning of certain psychological systems.

In spite of the physical 'togetherness' of the authors, the quality, completeness, and point of view of the various chapters vary widely. On the basis of obvious purpose and tone of writing, however, the chapters fall into two groups. One of these will be referred to as the 'handbook' chapters. The others do not really belong in a handbook.

The 12 genuine 'handbook' chapters make up more than half of the 800 pages of the volume. The following areas, with emphasis on abnormal behavior, are treated in different chapters: Psychomotor Functions; Expressive Movement; Sense Perception; Cognition; Heredity; Constitution; Childhood and Environment; Somatic Reactivity; Brain Damage; Psychosurgery; Electric Shock; and Drugs. Each of these chapters is an attempt at objective, informative coverage of significant work, including historical orientation, through about 1958.

The chapter on sense perception is a good review of the psychophysics of visual phenomena, and can be recommended to psychologists in all areas. The same can be said for the chapter on heredity. The chapter on cognitive abnormalities is an excellent review of a somewhat neglected area and includes a thorough analysis of concrete-abstract functioning in abnormal groups.

With respect to the other eight chapters (two of which are by the editor) it is clear to the reader that objectivity and informative coverage are not among the authors' primary intentions. The purposes are stated to be those of supporting a particular theory (without mention of rivals); to convince the reader of the adequacy of a particular method of doing research; and thirdly to express the private biases of the authors. These chapters include those on the following seven subjects: Clasification and Problems of Diagnosis; Motivation and 'Ego Functions'; Conditioning; Learning; Psychotherapy; Abnormal Animal Behavior; and Applied Abnormal Psychology. At times the style of writing in these chapters becomes so emotional and evaluative, that the reader can hardly be expected to have confidence in the adequacy of coverage.

Attempts at theoretical integration of factual material made in these chapters are primarily made in terms of three theories. The first of these is Eysenck's attempt to analyze abnormal behavior in terms of introversion-extroversion and the neurotic and psychotic factors. Yet nowhere in the book is there a concise presentation of this theory. Thorough acquaintance with this view is assumed. The second theory used for integration purposes is the Hullian Learning Theory. The third, and most adequately presented, of the 'theories' is the Yerkes-Dodson Law. While there should be a place in a handbook for a succinct presentation of theoretical views, an avowed attempt to integrate a wealth of material in terms of only a few theories, to the neglect of all the others, is unwarranted.

These chapters repeatedly express a bias in favor of factor analysis as the research method to be preferred in the area of abnormal behavior. The reader often gets the impression that the author thinks there is no other way of arriving at 'dimensions' than by factor analysis. It is not necessarily belittling this method to point out that its value is primarily one of preliminary identification of functional relationships. The trouble is that this approach produces a purely abstract region of dimensions, altogether divorced from concrete levels of behavior. The identification of a factor or dimension by the usual methods helps one very little at identifying the presence and degree of such a trait in a particular individual. The obvious need for following up factor analysis with a spelling out of specific operations for identifying the dimensions in a person's behavior is pointed out only once in these chapters (cf. p. 109).

The manner in which some of the authors' personal views are expressed is calculated to arouse in the reader mingled amusement, irritation, and indignation. For example, on p. 742, preliminary to discussing the work of Masserman on animals, the following sentence appears: "This approach is one to which the writer is unsympathetic; the reader is therefore cautioned against this bias." After reading this surprising statement several times, this reader was unable to decide whether the bias referred to was that of Masserman or of the writer of the chapter. Regardless of what one might think of the theoretical value of psychoanalysis, its undeniable importance in the area of abnormal behavior deserves more than this cavalier dismissal.

The chapter on psychotherapy is obviously written to convince the reader of Eysenck's private view that psychotherapy is of no value. The reader is left with the distinct impression that Eysenck is capable of very astute criticism when he does not like the results of a study; and that he can be astoundingly tolerant of defects in methodology when he does like the results. For example, although he spells out clearly and adequately the need of randomness in selecting control groups, he cites studies which have found psychotherapy to be of little value as ones in which control groups were used when they were not selected on a random basis. The only really adequate study that he presents is one concerned with delinquency and not with neuroticism or abnormal behavior.

One other chapter, which does not belong in a handbook, is the one on "Intellectual Abilities and Problem-Solving Behavior." The criticisms applied to the other seven, 'non-handbook' chapters does not apply to this one. This is a highly sophisticated analysis of the concept "test score." It is an excellent contribution to this body of literature. This chapter also contains an excellent critique of the conventional, testing-movement in psychology.

The present reader is not in the habit of looking for, or finding, misprints in books or articles. Some of the signs of poor proofreading in the chapters by the editor himself were, however, so glaring that they could not be passed over. On p. 16, for example, the wrong legend is printed under a large figure. In the chapters on drugs and psychotherapy, many of the references are not to be found in the bibliography; others occur with wrong dates.

It is unfortunate that this volume has chapters at the begnining and at the end which are most likely, by their language and treatment, to arouse annoyance in the reader. After an introductory statement of intention to select studies on the basis of an elegant functional design, the reader is likely to feel let down at the obvious inadequacy of some of the studies cited without criticism. When his annoyance has had time to subside, however, he will undoubtedly want to have this volume as part of his research library. The 12 'handbook' chapters, plus the wealth of material cited in some of the others, plus a good index, makes the volume a useful handbook for researchers in abnormal psychology.

VA Hospital, Jefferson Barracks, Mo. HENRY N. PETERS

Persuasian and Healing: A Comparative Study of Psychotherapy. By JEROME D. FRANK. Baltimore, The Johns Hopkins Press, 1961. Pp. xiv, 282, $5.50.

This book is the contribution of a man who combines in one person the highest attributes of scholar, scientist, and healer. It is a highly readable, deceptively simple book with a message which will bring glaringly into question many of the unsupported biases (foundations) upon which rest the teaching and practices of modern psychotherapy. The author, holder of a Ph.D. in Psychology from Harvard and an M.D. from the Harvard Medical School, is Professor of Psychiatry at The Johns Hopkins University and is a recognized authority on psychotherapy. It is his thesis that the field of psychotherapy is characterized by a multitude of conflicting theories and methods, but despite their apparent diversity all are attempts to heal through persuasion. As such, different approaches, including psychoanalysis, non-directive, eclectic, group, and other forms of psychotherapy, share features not only with each other but also with many other forms of persuasion and healing.

In a masterful introductory section, Dr. Frank reminds the reader that all societies at all times empower certain of their members to change the thinking and behavior of others in distress. By convincing argument, he points out that the witch doctor or *shaman* in primitive society, the *religious revivalist,* the psychotherapist, and even the recently introduced Communist *thought reformer* share this function. Also included in this group are the *training analyst* in a Psychoanalytic Institute, the psychologist conducting a *verbal conditioning* experiment, the physician who employs a *placebo* in his treatment of a patient, and, as an entity, the total group-multitude associated with religious cures (such as those observed at Lourdes). What are some of the factors common to effective influence in these fields? What implications are there in these common factors for the psychotherapist today, whose contribution is measured by his ability to help his patient change the attitudes and behavior responsible for his suffering?

Drawing on his own extensive clinical and research experience, and his wide familiarity with the scientific and cultural literature of our own and previous societies, the author ferrets out the common fundamental features of all persuasion and healing. What, for example, are the fundamental features of healing common to modern revivalism and primitive religions? He writes that, in modern religious revivalism, and in primitive societies, the *sufferer's* distress has a large emotional component, produced by environmental or bodily stresses, internal conflicts and confusion, and a sense of estrangement or isolation from his usual sources of group support. He tends to be fearful and despairing and to be hungry for supportive human contacts. The *persuader* (evangelist or shaman) and his group represent a comprehensive and pervasive world-view, which incorporates supremely powerful suprapersonal forces. That modern Communism identifies these forces with the Party, while religions regard them as supernatural is relatively unimportant for the purpose of this discussion, writes Frank. The world view is infallible and cannot be shaken by the sufferer's failure to change or improve. In return for submitting himself completely to a religious persuader and the powers he represents, the sufferer is offered the hope of surcease from suffering, resolution of conflicts, absolution of guilt and bodily pain, and warm re-acceptance by the group. The persuader is the point of interaction between the sufferer, his immediate group, and the suprapersonal powers. He guides the group's activities and embodies, transmits, interprets, and to some extent controls the suprapersonal forces. As a result the sufferer perceives him as possessing power over his welfare.

Similarly, writes Frank, in non-directive and psychoanalytic forms of psychotherapy, exploration of the patient's past is an effort to help him gain insight into his current maladjustments by uncovering their historical sources. From the emotional standpoint, detailed exploration of past history tends to mobilize guilt. In the absence of overt direction from the therapist, the patient is tempted to become ever more self-revealing, partly in the hope of producing something that will gain a supportive response from the therapist. Since the patient knows he is supposed to bring up troublesome and painful experiences, he does so, and many of these may make him feel guilty. Sooner or later the confessions and confidings begin to include material which the patient had not been aware of any need to confess and he soon gets into the position of 'betraying' himself and implicating others. He finds himself telling tales out of school and admitting things which he had previously denied—perhaps even to himself. So, whereas at first he had been relieved by the diminished pressure of his confessions, now new pressures develop because of them. Guilt tends to facilitate attitude change by alienating the patient from the experiences and attitudes that produce it. At the same time the psychotherapist's continued impartial interest in the face of guilt-arousing material tends to be experienced as a kind of implicit forgiveness. Since the therapist in some sense represents society, this may help to counteract the patient's sense of isolation and enable him to feel more closely integrated with his group. Thus the phenomena of guilt, confession, and absolution, which play such an important part in religious healing, conversion, and thought reform, have analogies in evocative psychotherapy. A related argument is made for directive psychotherapies.

The author, in his search for common features, then proceeds to include one of our major institutions for teaching psychotherapy, the psychoanalytic institute. In the learning of psychoanalytic theory and techniques in the psychoanalytic institutes, writes Frank, the *training analysis* of the psychoanalyst-intraining is probably the core of the learning (indoctrination) process. It consists of the application to the trainee of virtually the same technique as is used with patients. As with the sick patient being ministered by a shaman, or a prisoner undergoing thought reform, this technique involves a detailed review of the trainee's current thoughts and feelings in relation to his life history, interpreted consistently by the training analyst in terms of the doctrines of the Institute. Since the initiative in the production of material resides with the trainee, the method fosters on-the-spot improvisation and active participation on his part. As it lasts for several years, the same material is reviewed repeatedly. The training analysis is not complete until the student produces memories, thoughts, and feelings in a form that confirms the doctrines of the Institute. In developing this provocative thesis, as well as the earlier ones, Frank writes in a scholarly, philosophic vein, rather than in an argumentative tone. He is writing as a man willing to take his share of the responsibility for the present state of affairs.

This brief review can do littl emore than introduce the reader to some of Frank's broad scope and persuasive arguments regarding the place of psychotherapy in the contemporary and historical scene. In the opinion of this reviewer, Frank's view of psychotherapy will do much to place this human activity in its *important* but rightfully *minor* place in the affairs of man. The review of the pitifully small existing literature on what we know from research on psychotherapy, including Frank's own well-known studies on psychotherapy, is an excellent addition to the book and powerfully reinforces where we the practitioners and investigators of psychotherapy stand in our contribution to the welfare and knowledge of our fellow man.

University of Oregon Medical School JOSEPH D. MATARAZZO

The Etiology of Schizophrenia. Edited by DON D. JACKSON. New York, Basic Books, Inc., Pp. 456. 1960. $7.50.

What causes a young man just entering what should be the most productive period of his life to draw away from human company, to develop unpredictable and peculiar ways of thinking and acting, and to deteriorate into a state of helplessness? It is hard to imagine a greater mystery. Schizophrenia has been mysterious enough to challenge a wide range of scientists, from geneticists and chemists to psychologists and sociologists. Jackson, a Palo Alto psychiatrist of Sullivanian persuasions, has collected writings from these wide-ranging and often wrangling specialists and put them together in a useful and stimulating book.

The proposals about the origins of schizophrenia break down into two large opposing groups—the organic and the socio-psychological. Those who look for defects in the organism assert that the causes are genetic, constitutional, endocrine, biochemical, or neuropathic. Those who seek socio-psychological sources hypothesize sensory deprivation, early material rejection, confusing communication in the family, symbiotic attachment to an anxious parent, or the like. These opposing camps are echoes of the hereditary-environmental issue that has reverberated in many forms down through the centuries. Most of the authors in this book are willing to concede that schizophrenia may be a label covering a wide variety of differently caused conditions and that the condition is likely to result from an interaction between organic predispositions and learnings from a poor environment. Jackson, espousing a communications approach (p. 18), notes that the organic and psychological points of view toward research range from "observing the output with a nonintact computer" to "varying the input in an intact computer." Still there is enough venom in the attacks of the opposing scientists to make the controversy very interesting.

The geneticist, Book, reviews the work of Kallman and others as well as his own work with North Swedish population groups. His conclusion is that unless a "specific genetical prerequisite" exists, schizophrenia will not occur. In an extensive and scholarly rebuttal, Jackson casts doubt on the genetic argument, pointing out the many methodological problems, such as comparability of diagnoses, difficulty in separating environmental effects in the home from hereditary effects when one member of the family is schizophrenic.

Quoting Gerard's statement that there is "no twisted thought without a twisted molecule," Abood reviews biochemical research (p. 91). (Without denying the biochemical possibilities, I find this quotation highly misleading. It would be perfectly possible to learn mistaken premises and inadequate habits without having anything wrong with the organic substratum, just as a child in some cultures learns to fear all strangers.) Abood's extensive review supplemented by Kety's and Heath's more theoretical coverage points to the developing conceptual organization of many diverse empirical findings. Abood states, "It would appear that psychogenic disturbances are generally associated with an impairment in the action of acetylcholine, either through an interference with its action or availability" (p. 105), though he points out that "no one intrinsic metabolite or neurohumoral agent is outstanding in its relationship to the problem of psychoses." In another chapter, the story of the Zilla-x-notata is interestingly told by Bercel. The web patterns of these spiders become abnormal and rudimentary if they are injected with serum from catatonics but not from other schizophrenic types or normals. Sontag speculates on the possible relationship of maternal prenatal environment to the development of schizophrenia, but presents no directly pertinent studies. The effect of high anxiety or emotional arousal during gestation is of particular interest, since the fetal-maternal endocrine system largely interacts as a unit.

The psychological side is introduced by Windsor's long chapter (218 references). One theme which runs through much of Winder's excellent review is the importance of considering motivation in any study of schizophrenia. On several studies, if the patient's motivation is altered by rewards or need to escape an annoying stimulus, behavior improves to a point where differences from normals disappear. McReynolds is also interested in motivation. His chapter is perhaps the most stimulating one in the book. This new theory rests on the hypothesis that the person is engaged continuously in obtaining percepts and 'assimilating' or integrating them. He hypothesizes a natural tendency to seek new percepts and to 'perceptualize' them at an optimal rate. Unassimilated percepts, which are those which do not enter congruently into the person's organized conceptual system, accumulate and give rise to anxiety. Schizophrenia, then, is a result of an extremely high level of unassimilated percepts. The withdrawal of the schizophrenic is a selective avoidance made in order to reduce anxiety, but this withdrawal leads to a kind of functional sensory deprivation. McReynolds mentions several studies generated by this theory, but it is as yet in an early stage.

The last five chapters largely focus on the family orientation of the schizophrenic. Clausen and Kohn report on their sociological study of patients coming from Hagerstown, Maryland. Perhaps the most important finding was a pattern of strong maternal and weak paternal authority in families of schizophrenics as compared with controls. The three chapters by Lidz and Fleck, by Bowen, and by Weakland present hypotheses about family dynamics, all based on clinical observations of family interaction. Bowen sees the schizophrenogenic mother, not as rejecting the child, but as projecting her own inadequacy into the child and then mothering this inadequate part of herself. In this symbiotic relationship, the occurrence of anxiety in the mother becomes a signal to the child to help the mother by becoming her baby. Weakland elaborates on the now famous "double-bind" hypothesis of Bateson and his associates—the situation in which the patient is faced with two or more inconsistent messages in an intensive relationship. Schizophrenia is the resulting breakdown of communication. Weakland expands this concept to cover three-person interaction and administrative behavior. Several of these investigators have noted improvement in the schizophrenic patient as soon as the father and mother start to face their own conflicts frankly in family unit-therapy.

At the present time, there is no way of saying whether the approach of the organic or the socio-psychological proponents is better. Both directions need to be pursued, but the job of doing research in this area is almost as complex as all science. In conclusion, it can be said that Jackson's book is a good review and theoretical extension of earlier work on schizophrenia. It lacks an overall synthesis of these many diverse approaches, but this is too much to ask at this time. He could, however, have placed much more emphasis on basic processes of learning and on cross-cultural work. Boatman and Szurek conclude in the last chapter that their decade or more of intensive clinical work with schizophrenic children has been "too little and too late." For a wealthy society facing this most devasting attack on the human personality one cannot escape concluding that our research efforts have been too small and uncreative and the need is indeed urgent.

University of Oregon NORMAN D. SUNDBERG

The Aging American: An Introduction to Social Gerontology and Geriatrics. By MILTON L. BARRON. New York, Thomas Y. Crowell Company, 1961. Pp. xvii, 269. $7.50.

Because of the important social and cultural aspects of the problems of aging, Professor Barron believes that courses in gerontology will appear most often in departments of sociology and anthropology, psychology and social work. The present volume was prepared as an introductory course in social gerontology, with questions and research exercises at the end of each chapter, focused primarily on retirement and health problems of the aged in American society.

As the title implies, the author regards aging as a developmental process that continues throughout the life span. He points out that the rising interest in the aged, coupled with our youth-centered culture, has led to a neglect of research on the middle aged. Although seven of the fourteen chapters represent reprints of material originally published by Dr. Barron and his co-workers in various journals between 1953 and 1956, some of the chapters, as for example the one on Religion and Aging Adjustment, introduce new material.

The first two chapters of the book survey the development of gerontology in the United States and the problems of the aged. Three chapters in the second section attempt to develop a theoretical framework for sociological research in aging. The author pursues his thesis, first proposed in 1953, that the aged should be regarded as a quasi-minority group because they showed many of the characteristics of minority groups; *i.e.* they suffer from subordination, prejudice and discrimination with "group feelings of hypersensitivity, self-hatred and defensiveness," which is characteristic of minority groups. Barron regards the aged as a quasiminority group since they are not socially organized as independently functioning subgoups in American society but live as individuals within the very families of the alleged majorities. Although there is considerable evidence to substantiate the classification of the aged as a minority group, in the opinion of the reviewer the aged for whom problems are most acute are not usually those remaining within a family group.

Barron's thesis is that most of the problems of aging are man-made and are therefore controllable. Urbanization, the increasing gap between working life and life expectancy, discrimination against the employment of middle aged and elderly workers, and lack of formal education are identified as the causes of problems of the aging. From a sociological orientation, this point of view may be tenable, but nevertheless the deteriorative physiological and psychological events that occur during aging must also be taken into account in considering factors of aging. The author also points out that most norms which are regarded as desirable behavior for older people have been formulated in terms of middle class standards. Since, according to the author, most elderly people identify themselves with the 'working class,' the recommended activities such as hobbies, avocations, club activities and other nonvocational pastimes often have little appeal to the masses of the elderly. It is somewhat difficult to reconcile this statement with the growth and success of the activities of the Golden Age Club and other social groups organized for the elderly throughout the United States during the past few years.

A major portion of the book is devoted to detailed reports on a few selected sociological studies of the urban and rural aged. To the reviewer it would seem that in a textbook one would refer to some of the other important social surveys in the literature, as for example the classical Wolverhampton study by Joseph H. Sheldon, which is listed as a suggested reading but is not discussed in any detail. Although major emphasis is usually placed on the role of urbanization in producing problems for older people, the studies reported on rural populations show that the situation of the aged farmer is far from carefree. The strong orientation toward work, the attachment to the land, the problems attending the sale or transfer of the farm and a somewhat negative attitude toward activities of leisure time make the prospect of withdrawal from farming a difficult one for the farm operator.

The final chapters review some of the social services developed in the United States for dealing with problems of the aged, with a description of selected programs in Norway, Sweden, Denmark, Holland, and Belgium based on personal observations made by the author in 1954.

Throughout the volume many psychological characteristics are considered, such as attitudes, interests, anxiety, and personal characteristics of individuals, but few references are made to the psychological literature dealing with the effects of aging on these characteristics. The author implies that "they [the aged] demonstrate such typical minority group feelings as hypersensitivity, self-hatred, and defensiveness in reacting to their distinctive trait, old age." Without evidence of the relative frequency of these characteristics among the aged, the trap is set to fall into the stereotype categorization of the aged which the author rightfully decries.

The author is to be commended on his approach to aging as a developmental sequence that extends over the entire life span and his recognition of the importance of the middle years in contributing to behavior in the advanced years. All too often the senescent years are regarded as a special category standing in isolation from the total life span.

While the book has limitations in a psychological context, it will probably be of value as an ancillary source of information on some of the social aspects of gerontology.

Baltimore City Hospitals N. W. SHOCK

Psychoanalysis, Scientific Method, and Philosophy: A Symposium. Edited by SIDNEY HOOK, New York, New York University Press, 1959. Pp. 370, $5.00

We are told that "the contents of this volume comprise the Proceedings of the second annual New York University Institute of Philosophy," but less than, and more than, the proceedings of the Institute appear in the book. The first three of the four sections (Parts I-III) include papers given at the Institute (except for that of Dr. Kubie), but the papers of Part IV were submitted later. References to discussions occurring during the proceedings make one wish to know more about what was said at the meetings; for example, Professor Hook apparently made some interesting initial challenges to the assembly in his opening remarks. Another minor source of irritation is the inconsistent handling of references: some contributors give them, some do not; most include them at the end of their chapters, but at least one includes his references as he goes along; some are listed alphabetically and others in order of occurrence. Most important, however, as Professor Hooks says in his introduction, are the contributions made.

Dr. Heinz Hartmann begins Part I, "Psychoanalysis and Scientific Method," by referring to uses of psychoanalysis in philosophy for exploring problems from a new angle. He proceeds with a general view of psychoanalytic theory, based primarily on Freud, and points out problems of validating psychoanalytic hypotheses, stressing that validation can be best done within the analytic situation. He concludes with a sketch of experimental checks of psychoanalytic hypotheses. Professor Ernest Nagel follows with a clear set of criticisms of analytic theory and practice, emphasizing traditional criticisms. Dr. Lawrence S. Kubie discusses limitations of psychoanalytic method, including data-recording problems, controls and techniques of validation, and goes on to a discussion of psychoanalytic procedure in a clear, brief, and relatively non-technical statement, with emphasis on free association. He ends with a "philosophical statement" of function of symbolism in human life in its effects of time—and space—binding and of storing.

Part II "Psychoanalysis and Society," begins with a brief outline by Dr. Abram Kardiner of the histories of anthropology and of Freudian psychology. He finds each suffering from a faulty "frame of reference," the former because it "did not know how to include the human unit" and the latter because it "did not know how to include the environment" (p. 88). He concludes that "the autonomy of academic departments can be respected only at the peril of insuring our ignorance" (p. 90) and urges the need of empirical studies of primitive societies, of "our own troubled cultures" and of the restatement of recorded histories. Professor Van Den Haag disagrees with Dr. Kardiner's conclusions, finding them "irrelevant to any of the many actual and serious problems that anthropology and psychoanalysis seeks to solve" (p. 106), and says that Kardiner's own discussion has implicit two premises, "that psychological health is related to moral goodness" and "that society can be diagnosed almost as an individual patient can" (p. 111). Professor Alex Inkeles also takes off from Dr. Kardiner's analysis, agreeing "in general." He states further that personality psychology, not psychoanalytic theory, has been useful to sociological approaches, and that some reasons for the lack of application, or applicability, of the latter are its psychological reductionism, clinical overemphasis, lack of scientific method, overemphasis on instinct, interest in adjustment to the exclusion of role behavior, and so on.

Part III, "Psychoanalysis and Philosophy," is largely a wild goose chase so far as psychological readers are concerned because of Professor Morris Lazerowitz' flight into a literary sort of psychoanalytic speculation about some areas of philosophy. The chase of Professor Lazerowitz' conclusions, is joyfully take up by Professors Williams and Flew. Professor Lazerowitz' discussion is rather technical philosophically. He analyses various philosophical propositions and arguments and concludes that "psychoanalysis is relevant in a very special way to philosophical utterances. It alone can discover for us what they really say, as against what they delusively appear to say," (p. 153). Psychologists reading this book may find his discussion mainly, in a few of the many words of Professor Williams' rejoinder, "a semantic nuisance." Professor Donald C. Williams' contribution may be very little more relevant except to the interests of professional philosophers, but it has the virtue of being intelligible to the uninitiated, and witty. He disposes of Professor Lazerowitz in 20 pages, with an interspersed highlight of criticism of psychoanalysis in 2. Professor Antony Flew's analysis is also quite intelligible and of interest in itself.

Part IV includes 19 contributions, the larger part by philosophers, the selections varying in length from 1 to 25 pages, and varying also in interest and in relevance to the symposium. They are so varied that space will not permit an adequate summary here.

It is easy to conclude from this volume that interest in Freudian theory, or at least in discussing Freudian theory, is far from slackening; energy, joy and mild fury abound. It is perhaps a less easy conclusion, but a sneakingly suggestive one, that such collective efforts may not lead in themselves to the sorts of general and interrelated discussions and findings hoped for. An obvious fault of the conference is that proceedings were not organized to permit the discussions usefully to be included in the report. This is not an unusual fault with conferences but is one which can be lessened by some use of controlled or limited discussions, and discussants. A more serious fault is the slight use made of psychological and psychologist sources. Despite continual flurried calls for empirical checks and restatements of theory in experimental form, only a few of the contributors mentioned psychological literature of this sort. (Dr. Hartman gives more references to such sources than all the other contributors together.) Also, although it would be frivolous to suspect Professor Hook innocent regarding psychological materials and individuals, the only psychological participation in the symposium is in brief notes by Professors Kendler and Silverman. Finally, in the discussions of interdiciplinary overlapping, the weighting on the side of psychoanalytic interpreting of metaphysicians' personalities has made the result a little lopsided. Exploring possibilities of mutual help of metaphysics and psychoanalysis, whichever way it leans, can be at best of very general use and is most likely to result in "evanescent qualities" in a discussion (to use Professor Hook's expression). There are clearly areas where philosophers can be of some use to psychoanalysis, and Professor Scriven's chapter is an essay of this sort. Conceivably psychoanalysis might contribute more than a literary flavor to some areas of philosophy. Let us hope, with Professor Hook, that this volume will at least spark further interest in such a "community of inquiry."

Ithaca, New York ELISABETH C. LAMBERT

Organizational Behavior and Administration: Cases, Concepts and Research Findings. By PAUL R. LAWRENCE, JOSEPH C. BAILEY, ROBERT L. KATZ, JOHN A. SEILER, CHARLES D. ORTH, III, JAMES V. CLARK, LOUIS B. BARNES, and ARTHUR N. TURNER. Homewood, Illinois, The Dorsey Press, and Richard D. Irwin, 1961. Pp. xxv, 910. $7.95.

Eight members of the faculty of the Harvard School of Business have combined their efforts to produce a major textbook for the study of human relations in industry. The text consists of seven sections entitled "The Human Problems of Administration," "Work Group Behavior," "Supervisory Behavior," "Diagnosing and Proposing Remedial Action," "Intergroup Behavior," "Organizational Behavior and the Wider Culture," and "The Administrator as an Agent of Organizational Change." Each section consists of three parts: a short (two to four page) introduction, a selection of cases, and a collection of abstracts and readings from the literature which present "concepts and research findings." Some of the abstracts are just one page, while other readings are often reproduced in their original form. Section I, on "The Human Problems of Administration," has only one selection—an excerpt from Alexander Leighton's *Human Relations in a Changing World*—while Section II, on "Work Group Behavior," has no fewer than 19 selections from the literature. A total of 62 cases and 54 selections from the literature constitute the major portion of the book. Paul F. Lazarsfeld provides a foreword to the book in which he states that he is most impressed by the success of the publication in focusing on the relationship between the concrete experiences of the practitioner of human relations and the efforts of social scientists.

The book is addressed to those who intend to assume positions of leadership in business organizations. It attempts to expose the students to the realities of human social behavior, as well as to some of the major findings from hte social sciences that may integrate and improve their understanding of this behavior. Considering the state of our knowledge on how to teach skills in human relations, this is probably the most effective training device that can be employed at the moment. For students of industrial or industrial social psychology, however, the text has less value.

The case materials are not written to demonstrate a point about how people are supposed to act. They are written, instead, with careful attention to reproducing what was actually going on in the situation, no matter how confusing or contradictory some of the facts and events might seem. Such materials are very helpful as training devices for people who will have to make decisions as best they can. They are of less value to a social scientist who does not have to make such decisions as part of his work. The fundamental difference is that the scientist can afford to ask one question at a time, and so set up his experiments that the question can be answered and make gradual progress towards the understanding of a phenomenon. He tries to make statements that have a high probability of being true. The practitioner cannot afford to wait; the decision often needs to be made at once. Furthermore, if he makes decisions that are more often right than wrong, he is likely to be judged successful. Since the nature of the two jobs is so different, it is only reasonable that the training devices used for each student should be different.

Though the book was not written for students of the social sciences, the present reviewer thinks that students of industrial social psychology would find some uses for the present volume. First, the reading of the cases would provide them with vicarious experiences of "what is going on in industry" that they are not likely to get from their other books. Though these experiences are not likely to lead to systematic knowledge by themselves, they would be helpful in communicating a feeling about industrial reality. Since many students have not had the benefit of actual industrial experience, such descriptions of the industrial human environment might be useful to them. These cases have an advantage over novels and other literary works; they are not fictitious but, on the contrary, are attempts to reproduce events exactly as they occurred. Secondly, by studying a number of these cases the student is likely to extract some principles which he can check with his instructor, or the literature. This latter task is greatly facilitated by the inclusion of selections from the literature in the present volume.

This reviewer has the impression that the technique employed in this book is to use the case studies to create anxiety, arouse interest, etc. by the confusing state that characterizes human behavior in the raw, and then to use the selections from the literature to provide principles, concepts, and systems of thought about human behavior that will reduce this anxiety, will produce some confidence in the student, and will permit him to make decisions in the face of uncertainty. This is probably the best technique now known for the training of managers, but it has less merit for the training of social scientists. The latter would not benefit from so much certainty. In fact, the best attitude to cultivate in such students is one of "we know so little compared to what we do not know."

The selections from the literature emphasize the contributions by the members of the Harvard business school somewhat more than is desirable, and gloss over the inadequacies in design, and the methodological naïvete, that often characterize these studies. This latter aspect of the book makes it somewhat less desirable for unsophisticated students, who are likely to take the conclusions presented in these readings much more seriously than the sophisticated ones. If the instructor is willing however, to provide a critical evaluation of these studies for his students, then the book could be used with less sophisticated students.

Furthermore, the selections almost completely ignore laboratory studies in the group dynamics tradition that could shed light on the topics under discussion. Most of the studies reported in these selections have been done in field settings with whatever 'natural' controls were available, and with the exception of a few did not involve manipulation of variables by the researcher. The problem of coming to a firm conclusion about cause and effect relationships under these conditions is not sufficiently discussed.

In conclusion, this is an excellent text for the purpose for which it was designed

—to train practitioners in the art of human relations. For social science students, on the other hand, the text has much less value, though even such students could use it for certain specialized purposes, such as acquiring an acquaintance with 'industrial reality' (to the extent that this can be done through a book), and obtaining quick access to abstracts of many articles and books in the area of human relations.

University of Illinois HARRY C. TRIADIS

The Powerful Consumer: Psychological Studies of the American Economy. By GEORGE KATONA. New York, McGraw-Hill Book Company, Inc. 1960. Pp. lx, 276. $6.50.

This volume brings up to date the work of The Economic Behavior Program of the Survey Research Center at the University of Michigan, directed by George Katona since its beginning in 1946. Its major premise, as the title indicates, is that the consumer—at least in the American economy since 1946—has been a powerful force in determining the economic climate or health of our country rather than the passive victim of classical economic influences.

The book consists of four parts, plus a preface, appendix, bibliography and index. Part I consists of two chapters (30 pages) and is an excellent "Introduction." In it Katona points out some inadequacies of classical economic theory as applied to consumer spending, and then presents his point of view in two basic propositions: (1) "Demand depends on income and confidence," and (2) "Changes in confidence are measurable"—referring to consumers (p. 8). Confidence, he believes, is a function of consumers' attitudes toward their personal financial situation and toward the entire economy and their expectations of what will happen to it. Income is important to the extent that it provides for *discretionary* spending—primarily for durable goods such as automobiles, appliances, etc., and investments such as additions to and major repairs to homes.

Part II "Attitude Change," consists of Chapters 3-7, pp. 31-114. Chapter 3. "Consumer Attitudes and Demand for Durable Goods" discusses the economic fluctuations between 1945-1958. Chapter 4, "The Nature of Attitudes and Expectations," develops Katona's theory of the "economic psychology" of the consumer. Chapter 5 considers the origin of changes in attitudes, and discusses such psychological phenomena as "instability of response and acquisition of information," and "social learning." Chapter 6 is on "Attitudes and Behavior of Individuals." Chapter 7 deals with consumer saving and concludes that "it is the extent of discretionary spending by consumers rather than of savings which differs greatly between good and bad years; and it is in variations in such spending, strongly influenced by changes in attitudes, which may bring forth an upturn or a downturn in the economy" (p. 112). This part of the book would be more convincing if there were more evidence from surveys to show the extent to which this discretionary spending, as influenced by changes in attitudes, did produce *measurable and statistically significant upturns or downturns* in the economy.

Part III, "Psychological Findings," consists of Chapters 8-10, pp. 115-173. In this part Katona considers psychological influences as related to buying, *e.g.* "multiplicity of motives," "dissatisfactions and attractions," "levels of aspirations," and "group belonging." These are considered important agents which influence buying behavior.

Part IV, "Economic Fluctuations," consists of Chapters 11-14, pp. 173-247. In Chapter 11, "The Psychology of Prosperity," Katona points out that after the Second World War a majority of persons believed that another major recession such as we had in the 1930s could never happen again, primarily because the government could avoid one by activities which would employ large numbers of persons. Survey results showed high levels of satisfaction with occupational progress, with income, and with standards of living.

Chapter 12, "The Psychology of Inflation," tells us that "Inflation . . . is neither a purely financial nor a purely psychological phenomenon. It is the joint result of a gap between incomes and the supply of goods and of inflationary expectations" (p. 192). Basically, the majority of people were not worried about inflation.

Chapter 13, "The Psychology of Recession," analyzes the recession of 1957-58 from the points of view of the consumer and of the businessman. Survey data indicated that consumers were pretty well informed and that they were aware of what was happening in the economy. Perhaps the most important part of this chapter is the conclusion that "the maintenance of those long-range attitudes which characterized prosperity and the belief that the recession would be temporary, encouraged the consumers to spend rather than to add to their savings" (p. 222).

Chapter 14, "Consumer Sanity," presents the thesis that the consumer is not easily influenced to alter his economic behavior radically—either up or down. This conservatism enables him to stabilize the economy—again we find "The Powerful Consumer."

Much of the evidence of *how* powerful the consumer is might be called circumstantial or inferential rather than direct. The 'causal' relationship between attitudes and behavior is suggested by the survey findings but it is not demonstrated to have been responsible. Obviously, the basic premise that consumer attitudes and expectations influence buying behavior can be accepted as sound, and logically could be expected to have *some* effect on our economy. Two questions, however, arise: (1) How large is the influence of attitudes and expectations on buying behavior? (2) How large is the effect of this modification in buying behavior on our economy? This volume makes a good start toward answering these questions but it does not give us any precise quantitative information. It deserves careful reading, however, by the informed consumer, and by every social scientist interested in this area. An excellent bibliography cites related evidence as well as the basic papers and monographs of Katona and his colleagues.

Greenwich, Connecticut — ALBERT D. FREIBERG

Instinct, An Enduring Problem in Psychology: Selected Readings. Edited by ROBERT C. BIRNEY and RICHARD C. TEEVAN. Princeton, N. J., D. VanNostrand Co., 1961. Pp. x, 181. $1.45.

Reinforcement, An Enduring Problem in Psychology: Selected Readings. Edited by ROBERT C. BIRNEY and RICHARD C. TEEVAN. Princeton, N. J., D. VanNostrand Co., 1961. Pp. viii, 230. $1.45.

Color Vision, An Enduring Problem in Psychology: Selected Readings. Edited by

RICHARD C. TEEVAN and ROBERT C. BIRNEY. Princeton, N. J., D. VanNostrand Co., 1961. Pp. viii, 214. $1.45.

Evaluating a collection of readings is usually even more idiosyncratic than judging the merits of a systematic treatise. The wider the area covered by the selections, the more likely is difference of opinion over what should and should not be included. Probably much of the difference of opinion will involve questions of whether a particular topic should be included at all in a particular area, rather than the selection of papers to represent that topic. Thus there may be differences of opinion about whether to include the topic of instinct in a particular course, but we can agree pretty well on the most representative selections to include if that topic is to be presented. Birney and Teevan have therefore found a very effective solution to the problem of readings by developing these separate collections on specific topics which are usually subheadings in various courses.

Even within these specific topics there can still be differences of opinion about best selection of papers, especially if we ask specialists in the particular area. As a non-specialist, however, I can find little to criticize in their selection. In fact, I can think only of items I should like to add to the collections, and would want to add them without removing anything already present. The collection on "Instinct" begins with William James, and ranges to Frank Beach, and including Kuo, Lorenz, Hess, Tinbergen among the 14 articles. It is restricted to animal studies and avoids the discussions of instinct in man (James' chapter is cut to a very brief excerpt). "Reinforcement" is also strictly limited to animal studies, even Thorndike's selection has nothing to say about man. The editors have also omitted the Skinnerian kind of study of reinforcement, limiting the discussions to the pros and cons of reinforcement as an independently defined concept.

Since I hope that this series will be continued with collections on additional topics, I should like to close with some suggestions for improvement. There is evidence of hasty or careless editing in all three of these volumes. Some articles do not have a bibliography although there are references in the text (*e.g.,* Articles 10 and 13, in *Instinct,* 8 and 15 in *Reinforcement,* and 8 and 17 in *Color Vision*). Evidently when an article was excerpted or was part of a book with a general concluding bibliography, the editors did not take the trouble to extract the bibliography as well. Other evidences of haste include "XXX Southall" as the translator of Helmholtz's Physiological Optics (*Color Vision,* p. 10) and references to the "Psychology Review" (*Color Vision,* p. 113 and *Instinct,* p. 165).

Finally, the price for these paper-bound books seems high and could reduce their usefulness for an important function—supplementary reading for just one section or portion of a course. The price also seems high in view of the current practice of paying no royalties to the authors of articles reprinted in this manner. (If the publishers are changing this unfair practice, I shall be glad to withdraw my criticisms of the price). T. A. R.

BOOKS RECEIVED

(The books listed here have not as yet been noted in our pages.
Listing here does not, however, preclude their later review.)

ADAMS, ANDERSON, *How to Use Self-Hypnosis for Health and Wealth,* New York, Exposition Press, Inc., 1961, pp. 40, $2.50.

ADAMS, JAMES F., *Problems in Counseling,* New York, Macmillan Co., 1962, pp. 164, $2.25.

ADLER, GERHARD, *The Living Symbol. A Case Study in the Process of Individuation,* New York, Pantheon Books, Inc., 1961, pp. xii, 463, $6.00, Bollingen Series LXIII.

ADRIEL, JEANNE, *Soaring Sunward: Woman's Initiation,* New York, Pageant Press, 1961, pp. 242, $3.00.

ANANTARAMAN, VENKATRAMAN, *Mobility of Professional Economists in the United States,* Madison, Wisc., Industrial Relations Research Center, University of Wisconsin, 1961, pp. vii, 56.

APOSTEL, L., MAYS, W., MORF, A., and PIAGET, J., *Les Liaisons, Analytiques et Synthetiques dans les Comportements du Sujet,* Paris, Presses Universitaries de France, 1957, pp. 147, 600 f.

ARGYRIS, CHRIS, *Understanding Organizational Behavior,* Homewood, Ill., The Dorsey Press, Inc., 1961, pp. xii, 179, $5.00.

ARNOLD, MAGDA B., *Emotion and Personality,* 2 vols., New York, Columbia University Press, 1962, pp. vii, 296 and vii, 430, $7.50 each vol.

ARROW, KENNETH J., KARLIN, SAMUEL, and SUPPES, PATRICK, (eds.), *Mathematical Methods in the Social Sciences, 1959. Proceedings of the First Stanford Symposium,* Stanford, Calif., Stanford University Press, 1960, pp. vi, 365, $8.50.

BAHM, ARCHIE J., *Yoga: Union with the Ultimate,* New York, Frederick Ungar Publishing Co., 1961, pp. 162, $1.25.

BAKER, A., DAVIES, R. L., and SIVADON, P., *Psychiatric Services and Architecture,* Geneva, World Health Organization, 1959, pp. 59.

BARBANO, FILIPPO, *Partiti e Pubblica Opinione Nella Campagna Elettorale,* Torino, Instituto di Scienze Politiche dell 'Universita di Torino, 1961, pp. 298, Lire 2500.

BARTON, RUSSELL, *Institutional Neurosis,* Bristol, England, John Wright and Sons, Ltd., 1961, pp. 56, $2.50.

BATESON, GREGORY (ed.), *Perceval's Narrative. A Patient's Account of His Psychosis.* Stanford, Calif., Stanford University Press, 1961, pp. xxii, 331, $6.75.

BELLAK, LEOPOLD (ed.), *Contemporary European Psychiatry,* New York, Grove Press, Inc., 1961, pp. xxvi, 372, $3.95.

BELLER, E. KUNO, *Clinical Process,* New York, Free Press of Glencoe, Inc., 1962, pp. xx, 394, $10.00.

BENTLEY, F. MILDRED, *No, Johnny Can't Come Home,* New York, Vantage Press, 1962 (copyright 1961), 70, $2.50.

BERGLER, EDMUND, *Curable and Incurable Neurotics. Problems of "Neurotic" versus "Malignant" Psychic Masochism,* New York, Liveright Publishing Corp., 1961, pp. 471, $6.95.

BIEBER, IRVING and Associates, *Homosexuality,* New York, Basic Books, 1962, pp. viii, 358, $8.50.

BIRNBACH, MARTIN, *Neo-Freudian Social Philosophy,* Stanford, Calif., Stanford University Press, 1961, pp. vi, 283, $6.00.

BLAU, PETER M., and SCOTT, W. RICHARD, *Formal Organizations,* San Francisco, Chandler Publishing Co., 1962, pp. x, 312, $5.25.

BLEIER, RUTH, *The Hypothalamus of the Cat. A Cytoarchitectonic Atlas with Horsley-Clarke Coordinates,* Baltimore, Md., The Johns Hopkins Press, pp. 109, $15.00.

BLOS, PETER, *On Adolescence,* New York, The Free Press of Glencoe, Inc., 1962, pp. xii, 269, $5.00.

BORING, EDWIN G., *Psychologist at Large,* New York, Basic Books, 1961, pp. 371, $6.50.

BRAIN, RUSSELL, *Speech Disorders,* Washington, Butterworth, Inc., 1961, pp. v, 192, $8.50.

BRAZIER, MARY A. (ed.), *Brain and Behavior,* Vol. I, *Proceedings of the First Conference,* Washington, American Institute of Biological Sciences Publications Office, 1962, pp. 433, $7.75.

BRIGHTBILL, CHARLES K., *Man and Leisure,* Englewood Cliffs, N.J., Prentice-Hall, Inc., 1961, pp. xi, 292, $6.50.

BRITT, STEWART HENDERSON, *The Spenders,* New York, McGraw-Hill Book Co., Inc., 1960, pp. xiii, 293, $4.95.

BROSIN, HENRY W. (ed.), *Lectures on Experimental Psychiatry,* Pittsburgh, Pa., University of Pittsburgh Press, 1961, pp. vi, 361, $7.50.

BRUNER, JEROME S., *The Process of Education,* Cambridge, Mass., Harvard University Press, 1960, pp. xvi, 97, $2.75.

BRUSSEL, JAMES A., *The Layman's Guide to Psychiatry,* New York, Barnes and Noble, Inc., 1961, pp. xi, 235, $1.50.

BUSS, ARNOLD H., *The Psychology of Aggression,* New York, John Wiley, 1961, pp. x, 307.

CAMMELL, CHARLES R., *Aleister Crowley,* New Hyde Park, N.Y., University Books, Inc., 1962, pp. xxiii, 229, $6.00.

CANTONI, LOUIS J., and CANTONI, LUCILE, *Counseling Your Friends,* New York, The William-Frederick Press, 1961, pp. 105, $1.50.

CAPLAN, GERALD, *An Approach to Community Mental Health,* New York, Grune and Stratton, 1961, pp. ix, 262, $4.50.

CAPLAN, GERALD, (ed.), Prevention of Mental Disorders in Children, New York, Basic Books, Inc., 1961, pp. xii, 425, $8.50.

CAPRIO, FRANK S., and BRENNER, DONALD R. *Sexual Behavior: Psycho-Legal Aspects,* New York, The Citadel Press, 1961, pp. 384, $6.00.

CHAIX-RUY, J., *Psychologie Sociale et Sociometrie,* Paris, Librarie Armand Colin, 1960, pp. 210.

CHAPLIN, J. P., and KRAWIEC, T. S., *Systems and Theories of Psychology,* New York, Holt, Rinehart and Winston, 1960, pp. xv, 473.

CHOISY, MARYSE, *Psychoanalysis of the Prostitute,* New York, Philosophical Library, 1961, pp. 138, $4.75.

CLARK, KENNETH, *Vocational Interests of Nonprofessional Men,* Minneapolis, Minn., University of Minnesota Press, 1961, pp. xi, 129, $3.75.

COFER, CHARLES N., (ed.), *Verbal Learning and Verbal Behavior,* New York, McGraw-Hill Book Co., 1961, pp. vi, 241, $5.95.

COMFORT, ALEX, *Darwin and the Naked Lady,* New York, George Braziller, 1962, pp. xi, 174, $4.00.

COMMITTEE ON PSYCHIATRIC NURSING, *Toward Therapeutic Care. A Guide for Those Who Work with the Mentally Ill,* New York, Group for the Advancement of Psychiatry, 1961, pp. vi, 104, $2.50.

COWEN, EMORY, UNDERBERG, RITA, VERRILLO, RONALD, and BENHAM, FRANK, *Adjustment to Visual Disability in Adolescence,* New York, American Foundation for the Blind, 1961, pp. 239, $4.50 (cloth), $2.50 (paper).

CRAM, DAVID, *Explaining "Teaching Machines" and Programming,* San Francisco, Fearon Publishers, 1961, pp. vii, 84, $2.00.

CROW, LESTER D., and CROW, ALICE, *Child Development and Adjustment,* New York, The Macmillan Co., 1962, pp. xi, 514, $6.50.

DEESE, JAMES, LESLIE, LOUIS, and ZOUBEK, CHARLES, *Gregg Notehand,* New York, McGraw-Hill Book Co., Inc., 1961, p. 320, $5.95.

DE RIDDER, J. C., *The Personality of the Urban African in South Africa,* New York, The Humanities Press, Inc., 1961, pp. xvi, 180, $5.00.

DIACK, HUNTER, *Reading and the Psychology of Perception,* Nottingham, England, Peter Skinner Publishing, Ltd., 1960, pp. 155.

DREVER, JAMES, *Sourcebook in Psychology. A Course in Selected Reading by Authorities,* New York, Philosophical Library, 1960, pp. xxiv, 335, $6.00.

EDUCATIONAL TESTING SERVICE, *Invitational Conference on Testing Problems, Proceedings 1960,* Princeton, N.J., Educational Testing Service, 1961, pp. 136.

Engler, David, *How to Raise Your Child's IQ,* New York, Ballantine Books, 1961, pp. 153, 50¢.

Farber, Seymour, and Wilson, Roger (eds.), *Control of the Mind,* New York, McGraw-Hill Book Co., Inc., 1962, pp. xiv, 340, $6.50 (cloth), $2.95 (paper).

Finch, Bernard, *Passport to Paradise . . . ? ,* New York, Philosophical Library, Inc., 1960, pp. 191, $6.00.

Fischer, Hardi, and Trier, Uri P., *Das Verhältnis zwischen Deutschschweizer und Westschweizer,* Bern and Stuttgart, Verlag Hans Huber, 1962, pp. 82, Fr/DM 11.50.

Fonzi, Ada, *Problemi Di Psicologia,* Torino, Edizioni Universitarie, 1961, pp. viii, 247, L. 2500.

Foss, B. M. (ed.), *Determinants of Infant Behaviour,* New York, John Wiley and Sons, Inc., 1962, pp. xv, 308, $6.75.

Frame, Janet, *Faces in the Water,* New York, George Braziller, Inc., 1961, pp. 254, $4.00.

Gallaher, Art, *Plainville Fifteen Years Later,* New York, Columbia University Press, 1961, pp. xvi, 301, $5.00.

Garre, Walter J., *Basic Anxiety. A New Psycho-biological Concept.* New York, Philosophical Library, 1961, pp. 123, $5.00.

Geist, Harold, *The Etiology of Idiopathic Epilepsy,* New York, Exposition Press, 1962, pp. 297, $5.00.

Getzels, Jacob, and Jackson, Phillip, *Creativity and Intelligence,* New York, John Wiley and Sons, Inc., pp. xvii, 293, $6.50.

Goldfarb, William, *Childhood Schizophrenia,* Cambridge, Mass., Harvard University Press, 1962, pp. xxvi, 216, $4.50.

Green, Edward, *Judicial Attitudes in Sentencing,* New York, St. Martin's Press, 1962, pp. viii, 149, $5.75.

Greidanus, Johan H., *Fundamental Physical Theory and the Concept of Consciousness,* Oxford, Pergamon Press, 1961, pp. xi, 159, $5.00.

Grunspun, Haim, *Disturbios Psiquiatricos da Crianca,* Sao Paulo, Brazil, Fundo Editorial Procienx, 1961, pp. xxx, 576, Cr $1.450,00.

Hamilton, James A., *Postpartum Psychiatric Problems,* St. Louis, Mo., The C. V. Mosby Co., 1962, pp. 156, $6.85.

Harman, Harry H., *Modern Factor Analysis,* Chicago, University of Chicago Press, 1960, pp. 469, $10.00.

Havel, J. E., *La Condition de la Femme,* Paris, Librairie Armand Colin, 1961, pp. 221.

Henle, Mary (ed.), *Documents of Gestalt Psychology,* Berkeley, Calif., University of California Press, 1962, pp. xi, 352, $8.00.

Hendrick, Ives, *The Birth of an Institute,* Freeport, Maine, The Bond Wheelwright Co., 1961, pp. xiv, 164, $5.00.

Henrich, Edith, and Kriegel, Leonard, *Experiments in Survival,* New York, Association for the Aid of Crippled Children, 1961, pp. 199, $3.50.

Holland, James G., and Skinner, B. F., *The Analysis of Behavior,* New York, Mc-Graw-Hill Book Company, Inc., 1961, pp. x, 175, $3.50.

Huber, Jack T., *Report Writing in Psychology and Psychiatry,* New York, Harper and Bros., 1961, pp. x, 114, $3.50.

Hughes, Helen MacGill, (ed.), *The Fantastic Lodge,* Boston, Mass., Houghton Mifflin Co., 1961, pp. xiii, 267, $4.00.

Hunt, J. McV., *Intelligence and Experience,* New York, Ronald Press Co., 1961, pp. ix, 416, $8.00.

Jacobs, H. H., Suchman, E. A., Fox, B. H. *et al., Behavioral Approaches to Accident Research,* New York, Association for the Aid of Crippled Children, 1961, pp. viii, 178, $1.50.

Jaques, Elliott, *Equitable Payment,* New York, John Wiley and Sons, 1961, pp. 336, $6.00.

Jones, Marshall R., (ed.), *Nebraska Symposium on Motivation,* Lincoln, Nebr., University of Nebraska Press, 1961, pp. 210, $3.25 (paper), $4.25 (cloth).

Jung, C. G., *Collected Works,* edited by Read, Sir Herbert, Fordham, Michael, and

Adler, Gerhard, Vol. 4, Bollingen Series XX, New York, Pantheon Books, 1961, pp. xii, 376, $5.00.

Kallir, Alfred, *Sign and Design,* London, James Clarke and Co., Ltd., 1961, pp. 356, 55s net.

Keur, John Y., and Keur, Dorothy L., *Windward Children: A Study in Human Ecology of the Three Dutch Windward Islands in the Caribbean,* New York, Humanities Press, 1961, pp. x, 299, $6.50.

Klein, Josephine, *Working with Groups,* New York, Humanities Press, 1961, pp. 240, $7.00.

Klein, Melanie, *Narrative of a Child Analysis,* New York, Basic Books, Inc., 1961, pp. 496, $10.00.

Kremers, J., *Scientific Psychology and Naive Psychology,* Groningen, Holland, P. Noordhoff, 1960, pp. 132, $2.50.

Kuenzli, Alfred (ed.), *Reconstruction in Religion,* Boston, Mass., Beacon Press, pp. xvi, 253, $3.95.

Kuppers, Waltraut, *Zur Psychologie des Geschichts-unterrichts,* Bern, Verlag Hans Huber, 1961, pp. 157, Fr. 17.50.

La Barre, Weston, *They Shall Take Up Serpents: Psychology of the Southern Snake-Handling Cult,* Minneapolis, Minn., University of Minnesota Press, 1962, pp. 208, $3.75.

Lantis, Margaret, *Eskimo Children and Interpersonal Relationships,* Seattle, Wash., University of Washington Press, 1960, pp. xv, 215, $4.75.

Lasswell, Harold D., and Cleveland, Harian (eds.), *The Ethic of Power,* New York, Harper and Bros., 1962, pp. xvi, 509, $7.50.

LeBlanc, M., *Personnalité de la Femme Katangaise: Contribution a l'étude de son Acculturation,* Louvain, Editions Nauwelaerts, 1960, pp. 403, 380 FB.

Lee, Alfred McClung, and Lee, Elizabeth Briant, *Marriage and the Family,* New York, Barnes and Noble, Inc., 1961, pp. xxix, 367, $2.25.

Leuba, Clarence L. (in association with William F. John), *Man: A General Psychology,* New York, Holt, Rinehart and Winston, 1961, pp. x, 676, $7.95.

Levine, Edna, *The Psychology of Deafness,* New York, Columbia University Press, 1960, pp. xii, 383, $7.50.

Levinsohn, Florence, *What Teenagers Want to Know,* Chicago, Ill., Budlong Press Co., 1961, pp. 89, $1.50.

Levitt, Eugene, *Clinical Research Design and Analysis in the Behavioral Sciences,* Springfield, Ill., Charles C Thomas Co., 1961, pp. xxii, 199, $8.50.

Lewis, Don, *Quantitative Methods in Psychology,* New York, McGraw-Hill Book Co., Inc., 1960, pp. xii, 558, $9.50.

Locke, Norman, *Group Psychoanalysis,* New York, New York University Press, 1961, pp. ix, 253, $6.50.

Loosli-Usteri, M., *Praktisches Handbuch des Rorschach Tests,* Bern und Stuttgart, Verlag Hans Huber, 1961, pp. 187.

Lorand, Sandor, and Schneer, Henry (eds.), *Adolescents: Psychoanalytic Approach to Problems and Therapy,* Paul B. Hoeber, Inc., 1961, pp. xiv, 378, $8.50.

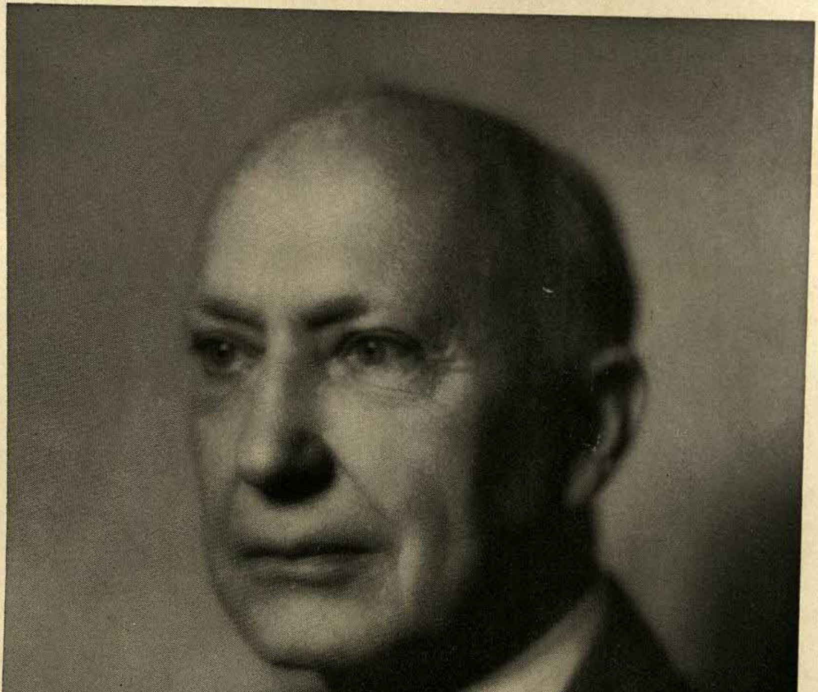
Mann, David, Woodward, Luther, and Joseph, Nathan, *Educating Expectant Parents,* New York, Visiting Nurse Service of New York, 1961, pp. viii, 223, $4.50.

May, Rollo (ed.), *Existential Psychology,* New York, Random House, 1961, pp. 126, 95¢.

McClintock, F. H., in collaboration with Walker, M. A., and Savill, N.C., *Attendance Centres. An Enquiry by the Cambridge Institute of Criminology,* New York, St. Martin's Press, Inc., 1961, pp. xiv, 152, $7.50.

McManus, William F., *Marriage Guide for Engaged Catholics,* New York, Paulist Press, 1961, pp. 128, $.75.

Meerloo, Joost A. M., *That Difficult Peace,* Great Neck, N.Y., Channel Press, 1961, pp. 190, $3.50.



Robert Sessions Woodworth

(See page 676)

THE AMERICAN
JOURNAL OF PSYCHOLOGY
Founded in 1887 by G. STANLEY HALL

Vol. LXXV DECEMBER, 1962 No. 4

LEARNING IN THE EARTHWORM, *LUMBRICUS TERRESTRIS*

By LOIS-ELLIN GREENE DATTA
Bryn Mawr College

About learning in annelids, very little is known. Yerkes was able to demonstrate maze-learning in the manure-worm as long ago as 1912,[1] and subsequent investigators (working with manure-worms and with earthworms) have confirmed Yerkes' finding, but the analysis of the phenomenon has not been carried very far. In the experiments to be reported here, the problem is taken up again as part of a broad program of research on the comparative psychology of learning.[2]

Yerkes' single *S* was given as many as 1,000 trials in a simple *T*-maze. The activating stimuli were light and touch; if the animal did not respond promptly to the light alone, it was stroked with a moist camel-hair brush. A correct turn at the choice-point permitted escape from the maze into an artificial burrow, while incorrect turns were punished (on early trials by contact with a strong solution of sodium chloride, and on later trials by shock from a pair of electrodes in the floor of the incorrect arm). Just before the electrodes, there was a strip of sandpaper "which was meant to serve as a warning against the electrical stimulus."[3] Although the behavior of the worm was rather erratic to begin with, improvement in performance soon appeared. There was an increasing readiness to enter the apparatus and to approach the choice-point, an increasing tendency to withdraw from the sandpaper, a reduction in retracing, and a rise in the proportion of errorless trips through the maze. Yerkes varied the number of trials per day and found some indication of better performance with 5 than with 10, 15, or 20 trials. The improvement in performance was relatively stable, persisting over a three-week period of no practice and after removal of the esophageal ganglia. When, after extended train-

---

[*] Received for publication April 28, 961. This research, reported in a Doctoral dissertation submitted to Bryn Mawr College, was directed by Professor M. E. Bitterman. It was supported by Contract Nonr 2829(01) with the Office of Naval Research and by Grant G-2246 from the National Science Foundation.

[1] R. M. Yerkes, The intelligence of earthworms, *J. anim. Behav.*, 2, 1912, 332-352.

[2] M. E. Bitterman, Toward a comparative psychology of learning, *Amer. Psychologist*, 15, 1960, 704-712.

[3] *Ibid.*, 351.

ing, the animal was required to go to the previously negative side, no significant indication of readjustment appeared in the course of 70 trials.

The study of maze-learning in annelids was taken up again in 1920 by Heck, who worked with several different species.[4] All the animals learned, the probability of a correct choice increasing progressively from a level slightly greater than chance to about 90% in 100-240 trials. Heck was able to distinguish several stages in the course of learning: generalized avoidance of the maze and of the choice-point; reduction in repetitive errors; withdrawal before shock but after partial entry into the incorrect arm; and increase in the probability of direct turns into the correct arm. Heck found more facility in habit-reversal than did Yerkes; 3 *Ss* trained to the 90% level of correct initial choice reversed in fewer trials than were required for the original learning. Heck, like Yerkes, found retention of the maze-habit after removal of the esophageal ganglia, and went on to show that the ganglia are not necessary for learning.

Maze-learning in annelids has been demonstrated in several more recent studies. Schmidt, comparing earthworms and manure-worms, found essentially equal rates of improvement, but what appeared to be a qualitative difference in the basis of improvement; manure-worms learned to withdraw from the sandpaper set before the electrodes, while earthworms did not.[5] Evidence of improvement in performance also has been reported by Arbit, who was interested in the relation between learning and the diurnal cycle, by Bharucha-Reid, who was interested in latent learning, and by Krivanek, who was interested in pseudoconditioning.[6] It should be noted, however, that not all attempts to demonstrate maze-learning in annelids have been successful. Swartz, using massed trials, trained worms of several species against preference in a *Y*-shaped tube after as many as 700 free-choice trials in which positional preferences were determined.[7] Incorrect choices were punished with severe shock, and correct choices were 'rewarded' by "picking up the tube and shaking the earthworm out of the fork, or pulling it out at the joint between the entrance tube and the *Y*-tube."[8] Under these conditions, training served only to decrease an initial preference for the incorrect side. Robinson's five earthworms showed strong initial positional preferences; trained against preference, they had succeeded only in achieving a chance level of performance when training was discontinued.[9]

About all that can be said with confidence on the basis of these studies is that the annelid is capable under certain conditions of significantly better than chance performance in a simple maze. There is some suggestion from experiments in which sandpaper was used that the sensory precursors of shock may come to 'signal' shock (that is, to evoke anticipatory withdrawal), although Robinson's failure to obtain like results with a plastic bar should be noted. Schmidt's work indicates that there may be some species differences here. There is some suggestion, too, that turning preferences can be modified—both Yerkes and Heck have reported increases in the proportion of direct turns into the correct arm of the maze (that is, of turns which are not preceded by probing into the incorrect arm), although it is possible that kinesthetic or other subtle sensory antecedents of incorrect turning, operating in the stem, may function in much the same manner as the sandpaper. Krivanek reports, however, that the extramaze stimulation of a given side of the animal can change turning preferences in the maze, a finding which suggests that an increase in the probability of correct turns is not necessarily due to learning about the maze. Clearly, more information is needed before the processes underlying the observed improvement in maze-performance can be understood. The conditions essential for improvement and the main variables which affect the course of learning have yet to be defined.

---

[4] L. Heck, Ueber die Bildung einer Assoziation beim Regenwurm auf Grund von Dressurversuchen, *Lotos Naturwiss.*, 68, 1920, 168-189.

[5] Hans Schmidt, Jr., Behavior of two species of worms in the same maze, *Science*, 121, 1955, 341-342.

[6] Jack Arbit, Diurnal cycles and learning in earthworms, *Science*, 126, 1957, 654-655; R. P. Bharucha-Reid, Latent learning in earthworms, *Science*, 123, 1956, 222; Jerome Krivanek, Maze-learning in annelids, *Physiol. Zool.*, 29, 1956, 241-250.

[7] R. D. Swartz, Modification of behavior in earthworms, *J. comp. Psychol.*, 9, 1929, 17-33.

[8] *Ibid.*, 22-24.

[9] J. S. Robinson, Stimulus substitution and response learning in the earthworm, *J. comp. physiol. Psychol.*, 46, 1953, 262-266.

METHOD

Subjects. The *S*s (246) were sexually mature specimens of *Lumbricus terrestris,* weighing 7-15 gm. They were selected from stock colonies which were maintained in covered 2-gal. tanks. The tanks were filled with a bedding mixture constituted of about ⅔ well-moistened, long-fibre spaghnum moss, ⅙ philodendron potting soil, and ⅙ shredded newspaper. The colonies were nourished by monthly sprinklings of fine-ground yellow corn-meal. Experimental *S*s lived individually in ½-pt. black plastic containers of the bedding mixture. All *S*s were stored in the dark, in a specially designed refrigerated chamber, with the temperature set at 10°C and the humidity controlled. Under these conditions, the animals seemed to thrive.

Apparatus. The maze shown in Fig. 1 was constructed of clear plastic and raised on 3-in. legs. Except for the length of the arms, which were shortened to reduce the time between response and its consequences (escape or shock), the dimensions of the maze approximated those used by Yerkes, Robinson, and Schmidt. The stem was 7½ in. long, ½ in. wide, and 1 in. high. A hinged plastic cover permitted easy insertion of the animal and prevented escape through the top. A plastic funnel, beginning 2½ in. before the choice-point and tapering to a centered ¼ × ¼ in. opening at the choice-point, was intended to aid orientation and to reduce retracing. The bar of the *T* was 4 in. long; each arm extended 1¾ in. from the ½ in. × ½ in. choice-area. The arms were covered by a hinged plastic lid and were grooved 1¼ in. from the choice-point to hold a sliding metal door which could be used to convert either arm into a cul. At the end of the correct arm, the worm's home-container was fitted, and the animal could continue directly from the maze to its burrow.

Two interchangeable floors were available for the bar of the *T*. One was made of the same clear, smooth plastic as the stem. The second (used in work on sensory control) was differentiated; the floor of one arm, from the end of the choice-area to exit (1¾ in.) was covered with small, smooth pebbles which were glued to the plastic. The removable floor slid into the maze in such a way that it was flush with the floor of the stem at the choice-point. It could be inserted with the gravelled surface in the right arm or in the left.

Both touch alone and light-plus-touch have been used as activating stimuli in studies of maze-learning in annelids. In all studies, incorrect turns have been punished by shock, and better performance seems to have been obtained with moderate (*e.g.* 15 ma.) than with intense (*e.g.* 300 ma.) shock. The experimental conditions used here were developed in pilot studies. While light alone and touch alone activated the worm, motivational control of the animal was better with a gradient of light and heat, supplemented by shock. A Westinghouse 12-v. automobile spotlight was suspended 6 in. above the stem of the maze, and a shadow-caster was used to produce the gradient. Five electrodes (¼ in. long) were inlaid in the stem and wired alternately to form a grid. With the grid, it was possible to shock animals for failure to leave the stem.

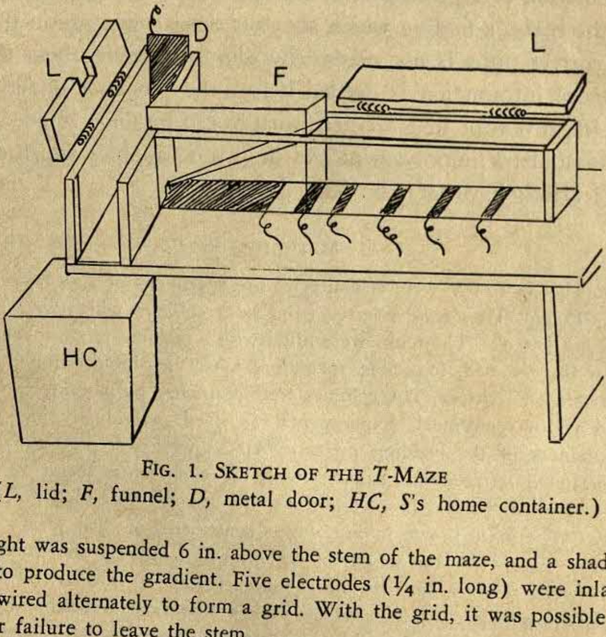


FIG. 1. SKETCH OF THE *T*-MAZE
(*L,* lid; *F,* funnel; *D,* metal door; *HC, S*'s home container.)

Shock also was used as punishment for error. The metal door which converted one arm of the maze into a cul was one of the electrodes which served this purpose, and the other was a large piece of metal (1¾ in. long) inlaid in the floor of the stem at the choice-point. Since the arms of the maze were short, the worm's body rested on the electrode at the choice-point while the animal was exploring the incorrect side, and contact with the metal door at the end of the cul completed the circuit. The shock for error caused a sharp withdrawal of the animal from the cul; the density of the current was greater at the anterior end of the animal, since the area of contact with the electrode at that end was smaller. The shocking circuit contained a sensitive relay which operated whenever current flowed through the animal and thus yielded a perfectly objective criterion of error.

Shock was provided by a circuit designed to produce constant current despite large variations in resistance. The line-voltage was stabilized with a voltage-regulator. A meter in circuit with the shock-generator gave a continuous check on shock-intensity.

Procedure. The same general procedure was used with all *S*s. The animal's individual container was removed from the refrigerator, the worm gently picked up by the middle segments, washed in cold water, and placed in the stem of the maze with the anterior segments toward the choice-point. If *S* did not move toward the choice-point of its own accord within 10 sec., a brief shock was administered from the floor of the stem. There was no other interference unless, as occasionally happened with large worms, the posterior segments were not withdrawn immediately into the home-container after a correct choice, in which case the floor-insert was tipped to slide in those segments. To prevent tracking, the maze was wiped thoroughly dry with fresh cottonwool after each trial; furthermore, each side was positive for half the *S*s, and the running orders were counterbalanced.

The time-limit for each trial was 180 sec., the longest time that worms could remain in the maze without danger of cumulative physical deterioration. An *S* that did not escape in 180 sec. was guided to the correct side with a cotton swab, and then left to complete the escape by itself.

The *S*s were trained in a room which was darkened except for the illumination provided by the light over the maze. In the early stages of the work, the temperature of the experimental room ranged from 21-27°C., although the temperature of the maze itself was far less variable because of the warming effect which the maze-light was intended to produce. All other observations were made in a subsequently completed, specialized laboratory with the temperature set at 21°C. Experimental observations were made throughout the day, but each *S* was worked with at approximately the same time each day.

The worms were given five trials per day. Yerkes, Heck, and Robinson also gave their *S*s about five trials per day, while Schmidt used 10 trials per day on alternate days. Except in an experiment on distribution of practice, the intertrial interval was 5 min. between trials, as well as between sessions, the *S*s were refrigerated. While Schmidt, Bharucha-Reid, and Arbit refrigerated their worms between sessions, the importance of cooling between trials has not heretofore been recognized.[10]

In all the experiments, with an exception to be noted later, the animals were trained at random, half to one side and half to the other. Free-choice trials have been used in the past to identify initial positional preferences, which neither Heck nor Swartz found to be marked or stable. Robinson's data do, however, give evidence of strong preferences. Pilot observations under the conditions employed here showed strong preferences only in a small percentage of the animals. Since the experiments also showed a tendency for free-choice trials to retard subsequent learning, it was decided to use random assignment to experimental conditions without preliminary tests.

SENSORY CONTROL

The very first experiments of this series were made with a relatively undifferentiated maze. Except for the metal door at the end of one arm (converting it to a cul) and the home-container at the end of the other, the two arms of the maze were similar in all respects but spatial location. In this situation, animals given the customary five trials per day, with an intertrial interval of 5 min., showed a familiar negatively accelerated learning curve; the percentage of correct initial choices rose to asymptote from the chance level in about five or six days. The only unique feature of the curve seemed to be its low asymptotic level, which was not very much greater than 70%. An undifferentiated maze was used in the pilot experiments out of an interest in examining the course of learning uncomplicated by the presence of obvious sensory precursors of shock, but now it seemed worthwhile to determine whether sensory enrichment would yield a higher asymptotic level of performance.

---

[10] Lancelot Hogben and R. L. Kirk, Studies in temperature regulation: I. The Pulmonata and Oligochaeta, *Proc. Roy. Soc., London,* 132, 1944, B, 239-252.

Three groups of earthworms, numbering 12 *Ss* each, were studied. The animals were selected at random from the stock colonies. For Group *SS*, the bar of the *T* had a smooth floor, as it did in the pilot experiments. Half the animals were trained with the right arm correct, and the others with the left arm correct. For the remaining two groups, *SG-C* and *SG-U*, the floor was differentiated: that of one arm was smooth and that of the other was gravelled. Half of the animals in Group *SG-C* were trained with gravel always on the left, and half with gravel always on the right; that is, tactual and spatial cues were confounded. Three *Ss* were trained to gravel-left, three to gravel-right, three to smooth-left, and three to smooth-right. For Group *SG-U*, tactual and spatial cues were unconfounded—the position of the gravel was varied systematically from trial to trial in accordance with Gellermann's rules[11]—and the problem became a pure tactual discrimination. The smooth arm was positive for half the *Ss*, and the gravelled arm was positive for the remaining *Ss*. All groups had 10 days of training with 5 trials per day and an intertrial interval of 5 min.

TABLE I

MEAN NUMBER OF CORRECT INITIAL CHOICES IN TEN DAYS OF TRAINING UNDER EACH OF THREE CONDITIONS*

Condition	SS	SG-C	SG-U
Combined	32.4	30.5	29.1
Smooth-positive	—	26.6	25.4
Gravel-positive	—	33.7	32.2

* The means are based on the data of 12 *Ss* in Group *SS* and 11 in each of the others (5 *Ss* in each of the smooth-positive subgroups and 6 *Ss* in each of the gravel-positive subgroups).

In Table I, the results of the experiment are summarized in terms of mean correct initial choice in the 50 training trials. Contrary to expectation, tactual differentiation did not improve performance, even in the confounded group. Although the between-groups variance was not statistically significant, there was, in fact, some indication of *poorer* performance in the differentiated maze on the part of the smooth-positive subgroups. As Table I shows, the gravel-positive *S*s performed as well as those of Group *SS* under both the confounded and the unconfounded conditions; learning curves plotted for Group *SS* and for the rough-positive subgroups closely resemble each other and those obtained in pilot experiments with the *SS* condition. By contrast, the performance of the smooth-positive *S*s remained about at the chance level. The effect of differentiation, then, was asymmetrical; the performance of the rough-positive subgroups was not improved, but that of the smooth-positive subgroups was impaired.

---

[11] L. W. Gellermann, Chance orders of alternating stimuli in visual discrimination experiments, *J. genet. Psychol.*, 42, 1933, 206-208.

Just why the smooth-positive subgroups did not learn, it is difficult to say without further investigation, and the problem of sensory control certainly is important enough to warrant intensive study. Nevertheless, the problem was not pursued further here. Instead, attention was focused on certain aspects of performance in the undifferentiated maze which became evident in the course of a more detailed analysis of the data. The analysis, which was prompted by the question of why the performance of the rough-positive subgroups was not better than that of the *SS* groups, revealed that the *SS* performance, in one sense, actually was very good already—much better than it had appeared to be on first inspection. The mediocre average daily scores turned out upon analysis to result from the combination of essentially random performance on the early trials of each day with relatively good performance on the later trials. New *S*s then were studied in the *SS* condition with a view to verifying this conclusion.

SOME GENERAL CHARACTERISTICS OF LEARNING IN THE UNDIFFERENTIATED MAZE

Results obtained with independent samples fit the generalizations suggested by the data of the experiments already considered.

The course of learning in the undifferentiated maze is displayed here in terms of the pooled data of several groups of earthworms, numbering 70 in all, which were studied under identical conditions in various subsequent experiments. There were 15 days of training, with five trials per day and a mean intertrial interval of 5 min. Half the *S*s, selected at random, were trained to go to the right, while the remaining animals were trained to go to the left.

In Fig. 2, the proportion of correct initial choice and the mean number of repetitive errors are plotted for each day of training. The curve for correct choice resembles the ones which were obtained in the experiments described previously. It is a negatively accelerated function rising from about chance on the first day to an asymptotic level which is achieved by about the fifth day. The low asymptote (approximately 70%) suggests that the animals are not capable of a high level of performance in the situation, but the incorrectness of this conclusion becomes evident when within-sessions curves are examined.

In Fig. 3, the proportion of correct initial choice is plotted as a function of trials for the first, fifth, and fifteenth days of training. All three curves begin near the chance level, that for the first day remaining at essentially the same level, and those for the subsequent days rising progressively throughout the session. Curves for the second, third, and fourth days would show intermediate rates of improvement within sessions, although with considerable overlapping of points, while curves for the days between the fifth and fifteenth would resemble those for the fifth and fifteenth days. In general, the curve for proportion of correct initial choice plotted in Fig. 2 reflects a change in slope over the family of within-sessions curves which is sampled in Fig. 3; that is, the between-sessions improvement may be accounted for entirely in terms of a change in the rate of within-sessions improvement. The mediocre asymptotic performance which appears in Fig. 2 is the result of combining relatively poor performance on the early trials of each session with relatively good performance on the later trials of the session. The fact that all of the within-sessions curves begin at about chance means that the specific preference for the correct turn which is ac-

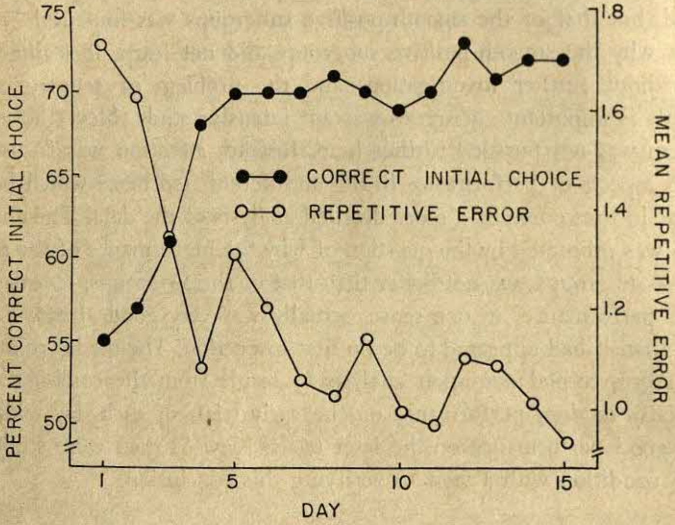


FIG. 2. MEAN PROBABILITY OF CORRECT INITIAL CHOICE AND MEAN REPETITIVE ERROR AS A FUNCTION OF DAYS OF TRAINING

quired within a session does not carry over from one session to the next. What *does* carry over is the *increased capacity to develop a specific preference* which seems to be acquired in the first three or four sessions.

The failure of a turning preference established within an experimental session to carry over to the next session has been reported by Yerkes, who found more than twice as many errors on the first trial of a session as on subsequent trials, even after as many as 495 training trials. Yerkes concluded at that point that there was "evidence of daily improvement" but not of "the acquisition of a habit which lasts from day to day."[12] In subsequent trials, with shock substituted for salt as punishment for error, and with the number of trials per day reduced to five, Yerkes began to find much better performance. There were virtually no errors after Trial 1 of each session, and, with further training, the animal might make as many as 20 consecutive errorless choices. These results reflect the development of a turning preference far more stable than was found in any of the experiments reported here. Heck also reported sequences of 19-20 errorless choices for all *Ss* after 100-240 training trials. Since there were 5-6 trials per day in Heck's experiment, the data suggest good retention of the acquired turning preference from one session to the next.

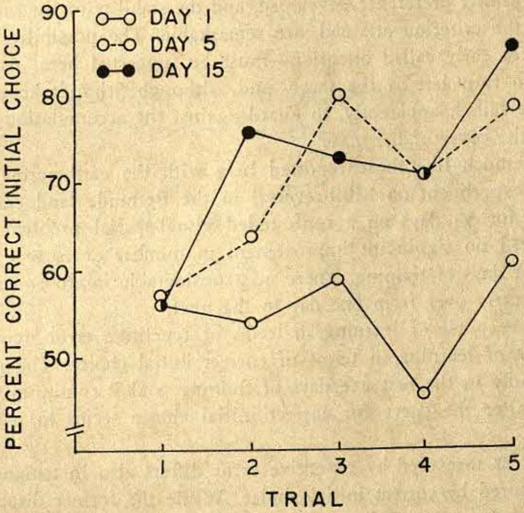


FIG. 3. MEAN PROBABILITY OF CORRECT INITIAL CHOICE AS A FUNCTION OF TRIALS ON THE FIRST, FIFTH, AND FIFTEENTH DAYS OF TRAINING

The discrepancy between the findings of Yerkes and Heck on the stability of acquired turning preferences and those reported here may perhaps be attributed to differences in amount of practice. In the present work, training never was carried beyond 24 days (120 trials). Although there was not the slightest bit of evidence of any sort of change in behavior between Day 5 and Day 24, the possibility must be considered that further training would have resulted in the development of more stable turning preferences. Another possibility is that the mazes employed by the earlier investigators were better differentiated; the attempt at differentiation of the present maze did not improve performance, but further efforts in this direction would appear to be in order.[13]

---

[12] Yerkes, *op. cit.*, 342.

The results of Schmidt are even more difficult to reconcile with the present results. Schmidt gave his animals 10 trials per day, every other day, in a differentiated maze. The first signs of learning—increased latency of response, backing out of the stem into the starting compartment, and withdrawal from junction-stimuli—did not appear until after 30–50 trials, yet a criterion of 10 out of 10 consecutive correct trials was achieved by 14 earthworms in an average of 60.5 trials. The rapidity with which a turning preference developed, and the stability of the preference which is implied by the criterion attained, are remarkable. The possibility of tracking—to which Yerkes early called attention—must be suggested here. All Schmidt's *Ss* were trained to turn left in the maze, and, although Schmidt kept changing the floor lining, he failed, apparently, to guard against the accumulation of mucous on the walls and the cover of the maze.

Results very much like those reported here with the earthworm were obtained in a previous experiment on habit-reversal in the Bermuda land crab.[14] A control group, trained for 33 days on a confounded visual-spatial problem in an escape-situation, showed no significant improvement in number of correct initial choices as a function of days of training. There was considerable improvement within days, but it did not carry over from one day to the next.

In Fig. 2, the course of learning in terms of repetitive error may be compared with the course of learning in terms of correct initial choice. The repetitive-error curve falls rapidly in the first five days of training, and it continues to fall in subsequent days, after the curve for correct initial choice seems to have reached its asymptote.

Improvement as measured by repetitive error differs also in temporal persistence from that measured by correct initial choice. While the former disappears entirely in the interval between sessions, the latter is more stable. In Fig. 4, the frequency of correct initial choice and the frequency of repetitive error are plotted for the first trial of each day of training. Frequency of correct choice remains at the initial (chance) level over the entire period, while frequency of repetitive error declines progressively in much the same manner as does the curve of repetitive error for all the trials of each day (which is plotted in Fig. 2). By Fisher's exact test, the difference in frequency of repetitive error in the first half as compared with the last half of training is significant at the 5% level, while the corresponding difference in correct initial choice does not approach significance.[15] In the study of the Bermuda land crab already noted, a parallel distinction between change in initial correct choice and in repetitive error was found. Mean repetitive error declined significantly over days, although the only improvement in correct initial choice was within days.

---

[13] It is interesting to note that Arbit, Bharucha-Reid, and Krivanek report about the same level of correct initial choice as that found here in about the same number of trials.

[14] L. G. Datta, Sandra Milstein, and M. E. Bitterman, Habit reversal in the crab, *J. comp. physiol. Psychol.*, 53, 1960, 275-278.

[15] R. A. Fisher, *Statistical Methods for Research Workers*, 1950, 96-97.

In summary, work with the undifferentiated maze reveals three kinds of improvement in the behavior of the earthworm. The first is an increase in the probability of correct initial choice. The second is an increase in the rate at which the probability of correct inital choice increases within sessions. The third is a decline in the frequency of repetitive error. The first kind of change is distinguished from the other two by its transience; it does not persist from one day to the next. The second and third kinds of change are distinguished in terms of their course of development; frequency of repetitive error continues to decline after the rate of within-sessions improvement in correct initial choice has stabilized. The functional relations among the three kinds of change remain to be established.

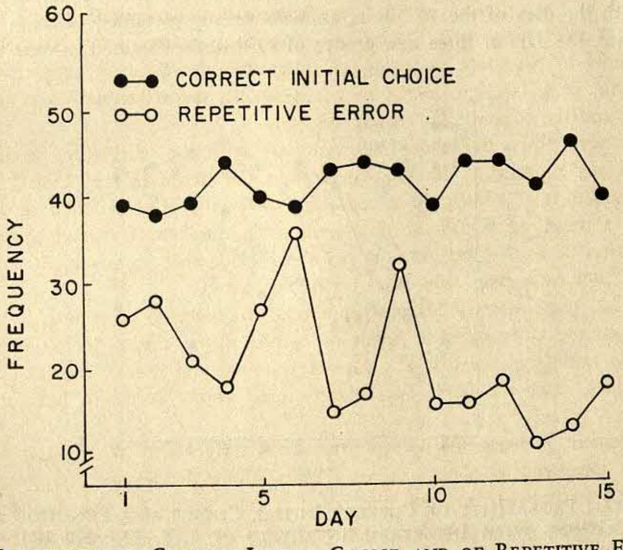


FIG. 4. FREQUENCY OF CORRECT INITIAL CHOICE AND OF REPETITIVE ERROR ON TRIAL 1 AS A FUNCTION OF DAYS OF TRAINING

DISTRIBUTION OF PRACTICE

In earlier studies of this series, a 5-min. interval between successive trials was used. This choice was based mainly on the implication in Yerkes' paper that an intertrial interval of 5 min. yielded better results than one of 20 min. Investigators since Yerkes have used intervals ranging from 0 min. (Swartz) to 20 min. (Schmidt) without apparent concern about the appropriateness of their choices. The data already reported here suggest, however, that distribution is a parameter of major importance, at least under the conditions of the present experiment.

Time between successive sessions is a special case of time between successive trials. Literally, it is not correct to say that a 5-min. intertrial was used in the earlier studies of this series. Two intertrial intervals actually were used: one of 5 min. (as between Trials 1 and 2), and one of about 23 hr. (as between Trials 5 and 6). The data of these studies show that a preference for the correct turn which is established in a series of relatively massed trials disappears entirely in the longer interval, and lead to the expectation that no preferences will develop if all trials are given at the rate of one per day. The data also suggest that decline in repetitive error may be more resistant to the effects of distribution than is increase in the probability of correct initial choice.

In the experiment now to be reported, the interval between sessions was unchanged, but the interval between the trials of each session was varied systematically. With the data of the 70 5-min. animals already presented taken as the point of reference, the data of three new groups of earthworms, trained as were the others in the undifferentiated maze, may be considered. For one group of 20 *S*s, the intertrial interval was one fifth as long as that of the reference group (1 min.); for a second group ($N = 21$), five times as long (25 min.); and for a third group ($N = 21$), 25 times as long (125 min.). The procedures employed for the new groups were the same in all other respects as those used for the reference animals. Half of each group, selected at random, was trained to the right, and half trained to the left. All animals, even those of the 1-min. group, were stored in the cold chamber during the intertrial interval.

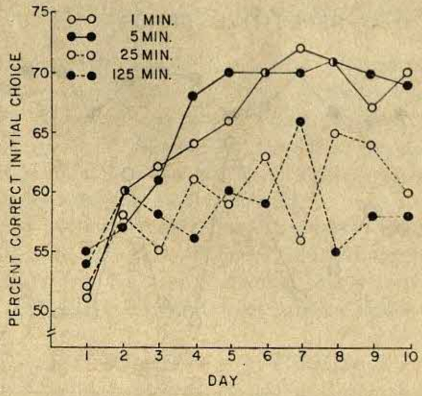


FIG. 5. MEAN PROBABILITY OF CORRECT INITIAL CHOICE AS A FUNCTION OF DAYS OF TRAINING WITH INTERTRIAL INTERVALS OF 1, 5, 25, AND 125 MIN.

The proportion of correct initial choice on each day is plotted in Fig. 5. The curves bear out the notion that distribution of practice is an important variable. While the performance of the 1-min. group is indistinguishable from that of the 5-min. group, the 25-min. and 125-min. groups are markedly inferior. These impressions are supported by the results of analysis of variance after (*arcsin*) transformation of the raw data. The analysis shows significant effects of Days and Groups ($p < 0.01$), although the interaction is not significant. The results of a separate analysis for the 25-min. and 125-min. groups give no indication at all of learning; the effect of Days falls far short of significance.

Within-sessions curves for the four groups, based on the pooled data of Days 6-10, are presented in Fig. 6. The curves reflect the relation noted earlier of between-sessions to within-sessions improvement. The 1-min. group, which shows the same degree of between-sessions improvement as the 5-min. group, also shows the same degree of within-sessions improvement. The 25-min. and 125-min. groups, which show no significant between-sessions improvement, also fail to improve within-sessions. Both sets of curves point to the conclusion that the change in the earthworm which is responsible for the increase in the probability of correct initial choice is quite unstable, lasting no longer than 25 min. Here, again, is a result which contradicts Schmidt, who found a higher level of performance than that of the present relatively massed groups with an intertrial interval of 20 min.; and again, an explanation in terms of tracking is suggested.

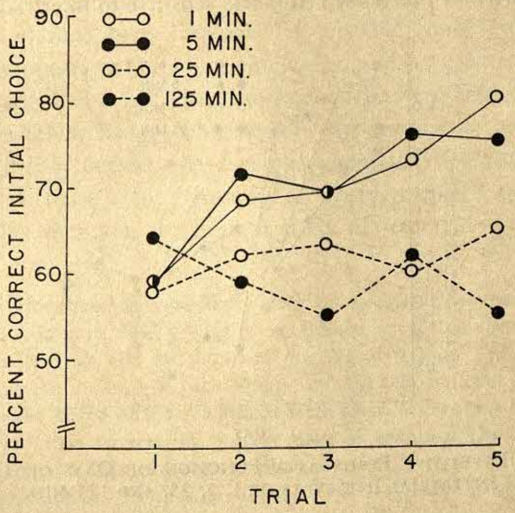


FIG. 6. MEAN PROBABILITY OF CORRECT INITIAL CHOICE AS A FUNCTION OF TRIALS IN THE LAST FIVE DAYS OF TRAINING WITH INTERTRIAL INTERVALS OF 1, 5, 25, AND 125 MIN.

It would be interesting to know whether any increase in the rate at which a turning preference develops (when trials are massed appropriately) results from training with trials so spaced that no preference develops. To find out, one would have to do a transfer experiment. For example, two groups might be trained for five days, one with a 5-min. interval and one with a 125-min. interval; then, for the next five days, the training of both groups would be continued with a 5-min. interval. It is conceivable that the two groups might do equally well in the second stage of the study, an outcome which would suggest that different processes underlie improvement in the capacity to develop a turning preference and the actual development of a preference. This question will be taken up again in connection with some experiments on habit-reversal.

While probability of correct initial choice reflects a process of learning which seems rather unstable over time, the data on repetitive errors point to the existence of a process which is less sensitive to intertrial interval. The

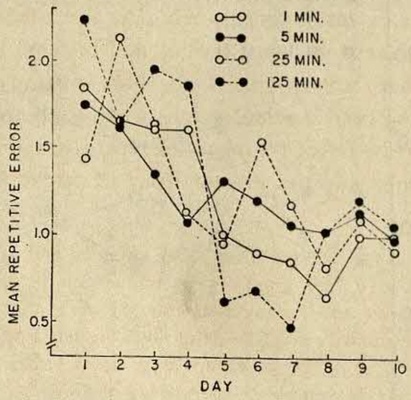


FIG. 7. MEAN REPETITIVE ERROR AS A FUNCTION OF DAYS OF TRAINING WITH INTERTRIAL INTERVALS OF 1, 5, 25, AND 125 MIN.

mean number of repetitive errors made by each group on each day is plotted in Fig. 7. All groups show essentially the same rate and degree of improvement. If anything, the curves suggest better elimination of repetitive error in the distributed groups, since these groups made more initial errors and thus had greater opportunity for repetitive error. It is possible, of course, to correct repetitive error-scores for frequency of initial error, but the statistical properties of the resulting index leave much to be desired. The best way to control for opportunity for error in the analysis of probability of repetitive error would be to control the occurrence of initial error. For example, the animal might be rewarded on some trials for whichever choice is made, while its first choice on other trials was punished. Such a procedure would prevent learning as measured in terms of correct initial choice, but it would permit a detailed analysis of the temporal course of the elimination of repetitive errors.

The main implications of these results on the effect of distribution are twofold: first, they emphasize the transient nature of the change in the animal which is responsible for improvement in the probability of correct initial choice; a 25-min. interval between successive trials is enough to prevent such improvement. Secondly, the results point to a distinction between changes responsible for increase in correct initial choice and those responsible for decrease in repetitive error; the two changes are differentially sensitive to the distribution of practice.

RETENTION

The preference for the correct side of the maze which is established under the conditions here employed is unstable; once established, the preference disappears entirely in a single day, and a measurable preference cannot be established at all when the interval between trials is 25 min. or more. The two other changes which occur—the increase in the rate at which a preference for the correct side develops when trials are massed appropriately, and the decrease in the frequency of repetitive error—are more stable. Just how stable they are, the present study of retention was designed to determine.

Sixty earthworms were selected from stock colonies and assigned at random to 6 groups of 10 *Ss* each. All groups were trained in the undifferentiated maze, with five trials per day and an intertrial interval of 5 min. The right side was positive for half the animals in each group, and the left side was positive for the remaining animals. Three of the six groups were given 5 days of training followed, after 1, 5, or 15 days of rest, by 5 more days of training. The remaining three groups were given 15 days of training followed, after 1, 5, or 15 days of rest, by 5 more days of training. The design, then, was a 2 x 3 affair, involving two degrees of original training and three retention-intervals.

In Fig. 8, the course of forgetting is plotted in terms of the proportion of correct initial choice on the five days of testing combined. The curves suggest a systematic decline in performance as a function of the interval between training and testing. Analysis of variance after (*arcsin*) transformation of the raw data shows the effect to be significant at the 1% level of confidence. The curves also suggest better performance after 5 than after 15 days of training, but the difference is not statistically significant. Continued training beyond the point at which the proportion of correct initial choice reaches asymptote apparently does not affect retention. Curves plotted for repetitive error rise over the 15-day interval in much the same manner as the curves for correct initial choice fall over the interval.

The performance of these animals in 5 days of testing is indistinguishable from their performance in the first 5 days of training when training and testing are separated by an interval of 15 days. Learning curves, both for correct initial choice and for repetitive error, are very much the same, and savings-scores based on both measures hover about zero. One may conclude, therefore, that even the more stable changes in behavior which are produced under the conditions of these experiments are unstable indeed by vertebrate standards. While these data are consistent with Krivanek's, it should be noted that Yerkes claimed much better retention for his manure-worm.[16] This finding, if indeed it is to be relied upon, may reflect either a species difference or a difference in amount or conditions of training.

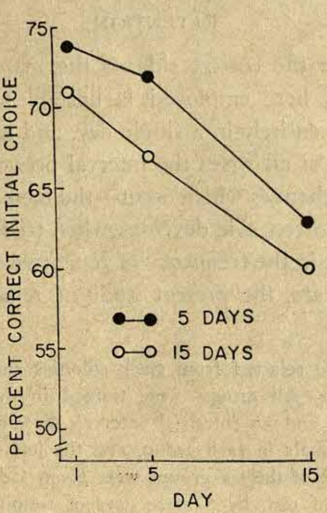


FIG. 8. RETENTION AS MEASURED BY RELEARNING AFTER 5 AND 15 DAYS OF TRAINING

HABIT-REVERSAL

Analysis of the data on probability of correct initial choice suggests a distinction between two different changes which occur under the conditions of these experiments. One is an increase in the probability of correct initial choice, and the second is an increase in the rate at which the probability of correct initial choice increases. The two changes differ markedly in stability over time, which suggests that they may be based on different underlying processes. One method of testing the hypothesis of independence already has been suggested in connection with the discussion of distribution. Another method, and the one actually employed, involves the study of habit-reversal. If the preference for a given turn which is established in a given experimental session disappears entirely in the interval between sessions, and if the improvement in the capacity to develop a preference in the course of a session develops independently of the preference which is established, it follows that performance should not be affected adversely by reversal, from one session to the next, of correct and incorrect turns. For example, after four days of training with left-positive, the animal should perform as well with right-positive as with left.

---

[16] "No indication of loss of the habit by the normal worm within three weeks appeared" (Yerkes, *op. cit.*, 349).

For the first experiment on habit-reversal, 10 earthworms were selected from the stock colonies. They were trained in the undifferentiated maze under the usual conditions (five trials a day with a 5-min. intertrial interval)—half to the right and half to the left. Positive and negative sides were reversed on the fifth day, on the ninth day, and on the thirteenth day, and training was terminated on the sixteenth day. That is, performance was studied over a series of four four-day problems, with one turn correct on Days 1-4 (Reversal 0), the opposed turn on Days 5-8 (Reversal 1), the original turn on Days 9-12 (Reversal 2), and the opposed turn on Days 13-16 (Reversal 3).

The results of this experiment are not consistent with the hypothesis of independence. Not only did performance in the first reversal fall below the level to be expected of unreversed animals on Days 5-8 (Fig. 2), but it was poorer even than performance on Days 1-4. The results suggest that turning preferences, once established, are after all carried over from session to session, but that they cannot be measured adequately in terms of correct initial choice on the early trials of a session. The random performance on Trial 1 may reflect, not the loss of preference between sessions, but the masking effect of some disturbing factor to which *S* must adjust anew in each session. Another possibility, which is suggested by the data on distribution of practice, is that the worm learns to respond to cues which are generated by experience in the maze and which fade in the intertrial interval; that is, the source of instability may be found in the discriminanda rather than in the learning.[17] It is interesting to note that the frequency of repetitive error did not increase in the first reversal, but continued to decline, and this fact must be taken as further evidence in support of the hypothesis that the processes responsible for improvement as measured by decline in repetitive error are different from those responsible for the improvement as measured by increase in correct initial choice.

---

[17] A previous experiment with the crab provides another case in which the persistence from session to session of a specific preference is revealed only by reversal. Unreversed controls showed no between-sessions improvement at all—only within-sessions improvement—yet their performance exceeded that of animals which were reversed periodically (Datta, Milstein, and Bitterman, *op. cit.,* 277).

The deleterious effect of reversal on the probability of correct initial choice did not, however, prove to be a permanent one. Performance in Reversal 2 was better than in Reversal 1, and performance in Reversal 3 was better than in Reversal 2. A nonparametric analysis of variance based on the data for Reversals 1-3 showed the improvement to be significant at the 2% level of confidence. As a further check on the reliability of this unexpected finding, another experiment on habit-reversal was performed.

A group of 20 *Ss* was trained for 4 days in the undifferentiated maze, half to the right and half to the left. There were 5 trials per day with an intertrial interval of 5 min. On the basis of their performance during this period, the animals were divided into two matched groups (Experimental and Control) of 10 *Ss* each. In the next 20 days, the Control Group continued to be trained to the originally positive side, while the Experimental Group was subjected to a series of five four-day reversals.

The results of this experiment were just what might have been expected on the basis of earlier findings. The Control Group behaved much like the animals whose data are summarized in Fig. 2. The performance of the Experimental Group (as measured in terms of initial correct choice) declined from Reversal 0 to Reversal 1, and rose progressively thereafter, until, by Reversal 5, it was doing *about as well as the Control Group*. Conventional analysis of variance (after *arcsin* transformation) did not, however, show the improvement in performance over the series of reversals to be statistically significant, and the inconsistency seemed traceable to a minority of animals who demonstrated marked positional preferences in Reversal 0.

For the purpose of providing a clearer picture of the course of habit-reversal in the earthworm, uncontaminated by strong initial preferences, a fresh sample of *Ss* was trained. Any animal which made fewer than three, or more than eight, correct initial choices on the first 10 trials (Days 1 and 2), or which failed to reach the level of 60% correct initial choice in Reversal 0 (Days 1-4), was discarded. Ten animals which could not be excluded on these grounds then were subjected to a series of five four-day reversals.

The performance of these animals, measured in terms of correct initial response, declined from Reversal 0 to Reversal 1, and rose progressively in subsequent reversals. Analysis of variance showed the improvement to be significant at a level of confidence well beyond 1%. The course if improvement is shown in Fig. 9, which is based on the data of the 10 *Ss* and the 6 *Ss* of the preceding experiment which met the same criteria for inclusion as did the *Ss* of the new sample. As the curve shows, the performance of reversed animals reaches the level of unreversed animals trained under the same conditions (see Fig. 2) after only about four reversals; thereafter, reversal is without significant effect.

At the mammalian level, of course, the phenomenon of progressive improvement in habit-reversal is well established, and the function plotted in Fig. 9 is much like those which have been obtained in analogous experiments with the rat.[18] What is surprising about the results for the earthworm is that convincing evidence of progressive improvement in habit-reversal has been rather difficult to find in experiments with simpler animals. The results of experiments with birds are contradictory.[19] A number of experiments with fish have given no evidence at all of progressive improvement, except in one instance, and there the pattern of improvement was qualitatively unlike that found in mammals.[20] At the level of arthropods, the data are contradictory and certainly provide no convincing demonstration of the progressive improvement which is so easy to find in mammals.[21] It seems reasonable, then, to view the results for the earthworm with a certain suspicion.

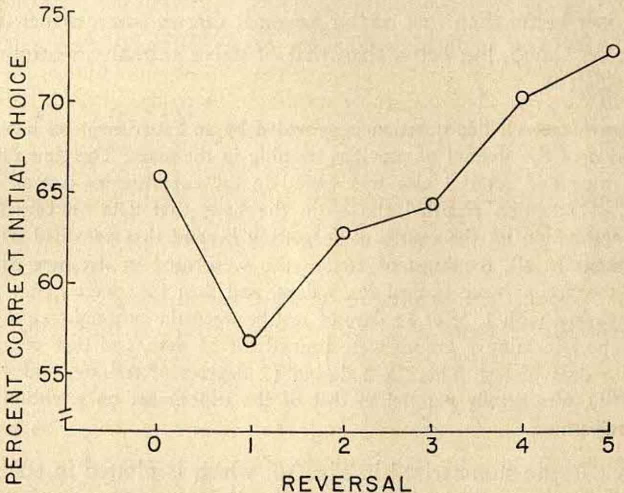


FIG. 9. THE COURSE OF HABIT-REVERSAL IN THE EARTHWORM

---

[18] I. Krechevsky, Antagonistic visual discrimination habits in the white rat, *J. comp. Psychol.,* 14, 1932, 263-277; F. P. Gatling, The effect of repeated stimulus-reversals on learning in the rat, *J. comp. physiol. Psychol.,* 45, 1952, 347-351.

[19] R. L. Reid, Discrimination-reversal learning in pigeons, *J. comp. physiol. Psychol.,* 51, 1958, 716-720; J. M. Warren, K. H. Brookshire, G. G. Ball, and D. V. Reynolds, Reversal learning by white leghorn chicks, *J. comp. physiol. Psychol.,* 53, 1960, 371-375.

[20] M. E. Bitterman, Jerome Wodinsky, and D. K. Candland, Some comparative psychology, this JOURNAL, 71, 1958, 94-110; J. M. Warren, Reversal learning by paradise fish (*Macropodus opercularis*), *J. comp. physiol. Psychol.,* 63, 1960, 376-378. The divergent findings are reported by Wodinsky and Bitterman, Discrimination-reversal in the fish, this JOURNAL, 70, 1957, 569-576.

[21] R. M. Yerkes and G. E. Huggins, Habit formation in the crawfish, *Cambarus affinis, Harvard psychol. Stud.,* 1, 1903, 565-577; T. C. Schneirla, Learning and orientation in ants, *Comp. Psychol. Monogr.,* 6, 1929, (No. 30), 1-143; Robert Thompson, Successive reversal of a position habit in an invertebrate, *Science,* 126, 1957, 163-164; Datta, Milstein, and Bitterman, *op. cit.,* 275-278.

The question may be asked, for example, whether progressive improvement in performance such as that shown in Fig. 9 must of necessity be interpreted as learning to reverse in the sense of a change which is due specifically to experience with reversal. Another possibility is that the improvement is due to increasing experience with the maze, and quite independent of reversal. This interpretation was suggested by the performance of the Control Group of the second experiment on reversal when it was reversed for the first time after 20 days of training. The performance was not only better than that of the Reversal Group when reversed for the first time on Day 5, but better than that of naive animals in original training on Days 1-4.

Further evidence on this question is provided by an experiment on habit reversal as a function of the amount of previous training in the maze. The time-interval between learning and reversal also was varied in the experiment—just as it was in the study of retention reported above—in the hope that data on reversal would provide information on the course of forgetting beyond that provided by the data on relearning. In all, 6 groups of 10 Ss each were used in the new experiment. Three of the groups were trained for 5 days, and then they were given 5 days of reversal training, with 1, 5, or 15 days of rest between the original training and the reversal. The remaining 3 groups were trained for 15 days, and then reversed after 1, 5, or 15 days of rest. The 2 x 3 design (2 degrees of training and 3 intervals of retention) was exactly parallel to that of the experiment on retention as measured by relearning.

The results are summarized in Fig. 10, which is plotted in terms of the proportion of correct initial choice for the five days of testing combined. The results for the relearning groups of the experiment on retention also are shown in order to provide a basis of comparison. Given five days of training, reversal and relearning groups differ markedly; analysis of variance after (*arcsin*) transformation shows the difference to be significant beyond the 1% level. Given 15 days of training, reversal and relearning groups differ much less markedly, and the difference does not approach statistical significance. Whether 5 or 15 days of training are given, the effects of training disappear almost entirely in 15 days of rest, and reversal groups, like relearning groups, perform very much as though they were experimentally naive. Reversal gives no more indication than does relearning, it seems, of any sort of retention over a 15-day interval.

These results suggest that the progressive improvement in habit reversal shown by earthworms subjected to a series of reversals should not be taken as evidence of 'learning to reverse,' although the basis of improvement remains to be discovered. At the mammalian level, a facilitating effect of overlearning on reversal has been reported by a number of investigators,[32] but attempts to find a like phenomenon in bird and fish have thus far been unsuccessful.[33] If there are indeed species differences in this respect, the results for the earthworm suggest that they are complexly determined. A possibility to be considered is that the facilitating effect on reversal is a perfectly general one, but that its operation in some animals is masked by especially powerful effects of differential reinforcement. Work on partial reinforcement in the fish provides an interesting case in point. For a time it appeared that the paradoxical effect of resistance to extinction, well known in mammals, was not to be found in the fish, but more recent experiments show the phenomenon and suggest that the earlier failure to

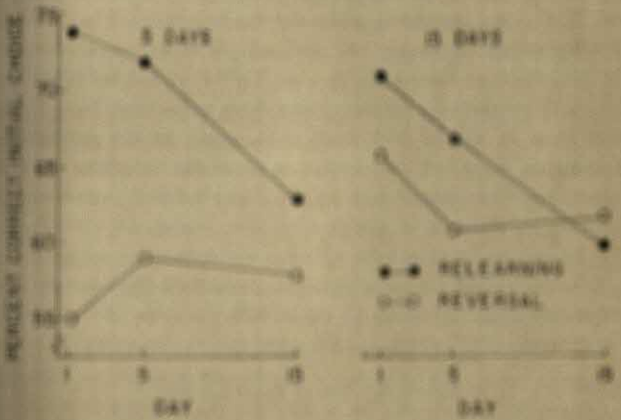


FIG. 10. RETENTION AS MEASURED BY RELEARNING AND BY REVERSAL AFTER FIVE AND 15 DAYS OF TRAINING

find it was due to the overriding effect on the fish of sheer frequency of reinforcement.[24]

---

[32] L. S. Reid, The development of noncontinuity behavior through continuity learning, *J. exp. Psychol.*, 46, 1953, 107-112; R. M. Pubols, The facilitation of visual and spatial discrimination reversal by overlearning, *J. comp. physiol. Psychol.*, 49, 1956, 243-248; E. J. Capaldi and M. W. Stevenson, Response reversal following different amounts of training, *ibid.*, 50, 1957, 195-198; David Birch, J. R. Ison, and S. E. Sperling, Reversal learning under single stimulus presentation, *J. exp. Psychol.*, 60, 1960, 36-40. In some mammalian experiments, the effect has failed to appear. See, for example, T. L. McCulloch and J. G. Pratt, A study of the presolution period in weight discrimination by white rats, *J. comp. Psychol.*, 18, 1934, 271-290.

[33] Warren, *op. cit.*, 570-578; Warren, Brookshire, Ball, and Reynolds, *op. cit.*, 573-575; Brookshire, Warren, and Ball, Reversal and transfer learning following overtraining in rat and chicken, *J. comp. physiol. Psychol.*, 54, 1961, 98-102.

Whatever its explanation, the facilitating effect of overtraining on reversal which is found in the earthworm has an important bearing on the question which prompted the study of reversal. That question concerns the relation between the processes underlying the increase (within sessions) in the probability of a correct turn in the maze, and the increase (between sessions) in the growth-rate of this probability. It was possible at the outset to entertain the hypothesis that the two changes are independent. Then the experiments which showed disruptive effects of reversal after small amounts of training seemed to contradict the hypothesis. The results suggested simply that relearning and reversal are more sensitive measures than is 'probability of correct response on Trail 1' of specific turning preferences which are carried over from one session to the next. Now the interpretation must be altered once more. The results of the experiment on overtraining do not affect the conclusion about the carry-over of turning preferences between sessions; such a carry-over must still be assumed. They do, however, point again to the development—independently of existing preferences—of an increased capacity for the acquisition of new preferences. Reversal provides a more sensitive measure of this capacity (which grows between Day 5 and Day 15) than does the slope of the within-sessions relearning curve (which reaches asymptote at about Day 5).

SUMMARY AND CONCLUSIONS

A series of experiments is reported on the behavior of earthworms in a simple T-maze, with light, heat, and shock as activating stimuli, with shock as punishment for incorrect choice, and with return to the home container as reward for correct choice. The effect of tactually differentiating the floor of the correct arm, or the floor of the cul, was studied in an early experiment of the series. The differentiation did not improve performance, and in all subsequent work an undifferentiated maze was used. Five trials per day were given, with an intertrial interval that was 5 sec. in all experiments but one in which distribution of practice was the independent variable. The measures of behavior were frequency of correct initial choice and frequency of repetitive error.

It appeared from the very outset that change in the probability of correct initial choice and change in the frequency of repetitive error might reflect the operation of different learning processes. The distinction is supported by the following results: (1) the learning curve plotted in terms of the first measure reaches asymptote sooner than does the curve plotted in terms of the second measure. (2) The two measures are differentially sensitive to the passage of time. Probability of correct initial choice grows within sessions and falls to chance in the interval between sessions; the learning curve plotted in terms of this measure rises over sessions only because the rate of within-sessions learning rises. Frequency of repetitive error shows no corresponding pattern of change. (3) Another indication of differential temporal stability is to be found in the fact that the spacing of trials to the extent of 25 min. prevents the development of a better-than-chance probability of correct choice but does not affect the course of repetitive-error elimination. (4) Frequency of repetitive error seems to be unaffected by reversal under conditions in which the probability of correct initial choice is affected quite markedly.

---

[24] R. C. Gonzalez, R. M. Eskin, and M. E. Bitterman, Extinction in the fish after partial and consistent reinforcement with number of reinforcements equated, *J. comp. physiol. Psychol.*, in press.

Measures of the probability of correct initial choice give evidence of two changes in behavior which are, at least to some extent, independent. One of these changes is the development of a preference for the correct arm of the maze. These lateral preferences were thought at first to be quite unstable (instability was suggested by the between-sessions decline to chance in the probability of correct initial choice and by results on distribution of practice), but experiments on habit-reversal provided evidence that (once established) they persist from one session to the next. The second change in behavior is an increase in the capacity to develop lateral preferences. Evidence for such a change comes from studies of habit-reversal, which show the same sort of progressive improvement that appears in analogous work with mammals. It happens, however, that experience with reversal is unnecessary for the improvement, which seems to develop as a function of the amount of training *per se*. A better mammalian analogue is the facilitation of reversal by overlearning.

All changes in the earthworm which are produced in these experiments are transient by mammalian standards. After 15 days of rest, the performance of a highly trained *S* is indistinguishable from that of an experimentally naïve one.

The relation of these findings to those of previous experiments on maze-learning in annelids is noted.

THE ROLE OF GRAPHEME-PHONEME CORRESPONDENCE IN THE PERCEPTION OF WORDS

By ELEANOR J. GIBSON, ANNE PICK, HARRY OSSER, and MARCIA HAMMOND, Cornell University

What is the critical unit of language for the reading process—the perception of written words? Recommendations for the teaching of reading have usually depended on the choice and definition of a unit. It was once assumed that the letters of the alphabet were the units, so children began reading instruction by learning their ABCs. The advent of Gestalt psychology (applied in a very simple-minded way) changed all this. Several generations of children, as a result, were taught by the 'whole word' method. As Bloomfield points out, such a method loses all the advantages of an alphabetical language.[1] One might as well learn picture-writing. Yet the letters of our alphabet in many cases do not have a one-to-one correlation with phonemic units. Is there, then, a proper unit?

It is the purpose of this paper to investigate the hypothesis that a unit is constituted by spelling-to-sound correlations. This is a higher-order unit formed by grapheme-phoneme correspondences. It is not defined by speech alone, or writing alone, nor is it a morpheme. It is the letter-group which has an invariant relationship with a phonemic pattern. Whole words usually have such a relationship; but often they can be broken into smaller clusters of letters which still have the kind of relationship referred to when they are in a stated position relative to other such clusters. The clusters may be of different sizes and the rules for the grapheme-phoneme correspondence are conditional on what precedes or what follows. The correspondence is not a matter of 'sequential probability' but it is to be found in the structure of written English *as it is related to spoken English.*

The existence of such correspondence has been demonstrated in a study by Charles F. Hockett and a group of co-workers in the Department of Linguistics at Cornell University. This group has classified English graphic monosyllables (1) according to the arrangements of letters in them, and (2) in terms of their pronunciation. The aim of the classification is to discover the rules by which pronunciation can be predicted from spelling. A graphic monosyllable includes a vowel-spelling, an initial-consonant

---

* Received for publication August 1, 1961. This research is part of an interdepartmental project for study of the psychological processes basic to reading, supported by a grant from the U. S. Office of Education.

[1] Leonard Bloomfield, Linguistics and reading, *Elementary Engl. Rev.,* 19, 1942, 125-130, 183-186.

spelling, and a final-consonant spelling. An initial-consonant spelling might be *CL*; a vowel-spelling, *EA*; a final consonant spelling, *TS*. The monosyllable is *CLEATS*. Rules for pronunciation are formulated conditionally on what precedes and follows as well as in terms of the letter-group as such. Regularities in spelling-to-sound predictability may be termed spelling-to-sound correlations, or grapheme-phoneme correspondences.

A child who learns to read (well) is forming useful spelling-to-sound habits based on these rules, whether he could tell you so or not. Even if he is taught by some method which makes it difficult, he must eventually discover the important spelling-to-sound correlations, if he is to be able to generate for himself the way to read new words.

From this starting position, we can proceed to the statement of a psychological hypothesis which has some testable consequences. Reading consists of decoding graphic material to the phonemic patterns of spoken language which have already been mastered when reading is begun. The units to be decoded are not single letters, for these have no invariant acoustic match in our language. The whole word is possible, but is uneconomical as a training unit for it provides no basis for independent decoding of new graphic combinations. The hypothesis advanced is that the reading task is essentially that of discovering higher-order invariants, the spelling-to-sound correlations. These are constants which are presumably discovered by exposure to both the graphic and the phonemic stimuli at the same time and in different contexts, so that the invariant combinations can be recognized in many different words. There are not a fixed number of units, such as all the words in the dictionary. There are, rather, all the potential ones which the rules of correspondence generate.

It is assumed that the individual discovers these grapheme-phoneme correspondences as he learns to read, even if he is not specifically taught them. He may not be able to draw up a set of rules but he 'has' them, if he is a good reader, just as a young child 'has' grammar long before he can formulate rules for it. Once he has them, any letter-combinations which follow the rules are functional units.

From this hypothesis we predict that a skilled reader should discriminate better visually, letter-groups (pseudo-words) which are constructed according to the rules of spelling-to-sound correlation (found in the structure of written and spoken English) than ones which are not, or are only partially so. This hypothesis is tested in the experiments described below. Recognition of letter-groups presented tachistoscopically was the criterion chosen for discrimination. Ability to perceive a letter-group presented tachistoscopically was found by Crausman to correlate with reading ability.[2] There is reason to think, therefore, that it is a valid measure of word-discriminability in the reading process.

HISTORICAL BACKGROUND

There is a certain amount of relevant material in the literature of experimental psychology which it may repay us to examine in the light of this hypothesis. Particularly, a number of experiments on word- and letter-recognition have been conducted in recent years with the method of tachistoscopic viewing. Howes and Solomon said, in reporting one of these, "any general property of words . . . should be operative in the word-threshold experiment, and the dependent variable (duration threshold) some function of it."[3] What is the logic of this statement, and what are the properties of words (or letter groups) which have been treated as parameters in a 'word-threshold' experiment?

The logic is discussed by Solomon and Howes. Perception, they say, is defined by both stimulus-properties and response-properties. When linguistic responses are elicited in a perceptual experiment, "any variable that is a general property of linguistic responses must also be a property of any perceptual concept that is based upon these responses."[4] *Word-frequency* was the property investigated by Solomon and Howes. They defined it simply in terms of frequency in the Thorndike-Lorge tables, but pointed out that frequency could refer to "frequency of visual exposure of a linguistic *stimulus*" or "the frequency of occurrence of a linguistic *response*." The two are different, they thought, but probably correlated. The assumption has been expressed by Postman and Conger that *emitted* frequency of a word, as such, affects the visual recognitive threshold.[5] Osgood, on the other hand, questions this assumption: "frequency-of-conversational-usage would not affect recognizability of words as *visual* stimuli."[6]

Postman and Rosenzweig found, in line with Osgood's position, that auditory training (hearing and speaking) did not significantly decrease visual recognitive thresholds for printed syllables, although it did decrease auditory thresholds, and visual training decreased visual thresholds.[7] Sprague also found that oral practice alone had no effect on tachistoscopic recognitive thresholds.[8] On the other hand, Forrest found that visual thresholds (determined by increasing sharpness of focus) varied with relative familiarity of paralogs, although the method of familiarizing was by auditory presentation.[9] (It should be noted that the paralogs were all 'pronounceable' and that they were exhibited twice in written form so that *S* would be familiar with the spelling of the words.)

---

[2] Burt Crausman, Temporal perception of good and poor readers, *Canad. Psychol.*, 1, 1960, 113 (abstract).

[3] D. H. Howes and R. L. Solomon, Visual duration threshold as a function of word probability, *J. exp. Psychol.*, 41, 1951, 401.

[4] R. L. Solomon and D. H. Howes, Word frequency, personal values, and visual duration thresholds, *Psychol. Rev.*, 58, 1951, 257.

[5] Leo Postman and Beverly Conger, Verbal habits and the visual recognition of words, *Science*, 19, 1954, 671-673.

[6] C. E. Osgood, *Method and Theory in Experimental Psychology*, 1953, 293.

[7] Leo Postman and M. R. Rosenzweig, Practice and transfer in the visual and auditory recognition of verbal stimuli, this JOURNAL, 69, 1956, 209-226.

[8] R. L. Sprague, Effects of differential training on tachistoscopic recognition thresholds, *J. exp. Psychol.*, 58, 1959, 227-231.

The hypothesis proposed in this paper attempts to cut across the dichotomy of stimulus- and response-frequency by proposing that the critical unit to be considered in word-recognition involves both—that it is, in fact, a stimulus-response *correspondence* which the individual becomes skilled in detecting as it recurs in an invariant relationship in different words. Neither frequency of visual exposure alone or emitting the oral response alone is of significance. Rather, it is frequency of experiencing a grapheme-phoneme coincidence which leads to skilled recognition.

No experiments have been performed heretofore with this concept in mind, but the experiments which have attempted to relate some property of words or letter-groups to visual discriminability will be briefly summarized.

Word-frequency. The hypothesis that greater frequency of words in the language should be correlated with speed and accuracy of recognition has been verified in a number of experiments. Cattell's early experiments with groups of letters and words exposed on a revolving drum laid the foundation for this research,[10] followed by Erdmann and Dodge.[11] Vernon,[12] Howes and Solomon,[13] Solomon and Postman,[14] McGinnies, Comer and Lacey,[15] Postman and Conger,[16] King-Ellison and Jenkins,[17] Baker and Feldman,[18] and Moscovici and Humbert[19] have all confirmed the role of word-frequency (in some cases built in experimentally) in lowering visual recognitive thresholds.

Experimental findings for 'familiarity' of words are comparable. Using Noble's technique of scaling words for familiarity by means of a rating scale,[20] Kristofferson obtained recognition-thresholds for 20 items, some real and some pseudo-words, and found a significant correlation between rated familiarity and recognition-threshold.[21]

---

[9] D. W. Forrest, Auditory familiarity as a determinant of visual threshold, this JOURNAL, 70, 1957, 634-636.

[10] J. McK. Cattell, Ueber der Zeit der Erkennung und Benennung von Schriftzeichen, Bildern, und Farben, *Philos. Stud.*, 2, 1885, 635-650.

[11] Benno Erdmann and Raymond Dodge, *Psychologische Untersuchungen über das Lesen auf experimenteller Grundlage,* 1898, 360 pp.

[12] M. D. Vernon, *Studies in the Psychology of Reading: A. The Errors Made in Reading, Med. Res. Council, Spec. Rep. Ser.*, No. 130, 1-40.

[13] Howes and Solomon, 1951, *op. cit.*, 401-410.

[14] R. L. Solomon and Leo Postman, Frequency of usage as a determinant of recognition thresholds for words, *J. exp. Psychol.*, 13, 1952, 195-201.

[15] E. M. McGinnies, P. B. Comer, and O. L. Lacey, Visual recognition thresholds as a function of word length and word frequency, *ibid.*, 44, 1952, 65-69.

[16] Postman and Conger, *op. cit.*, 671-673.

[17] Patricia King-Ellison and J. J. Jenkins, The durational threshold of visual recognition as a function of word frequency, this JOURNAL, 67, 1954, 700-703.

[18] K. E. Baker and Herman Feldman, Threshold-luminance for recognition in relation to frequency of prior exposure, this JOURNAL, 69, 1956, 278-280.

[19] S. Moscovici and T. C. Humbert, Usage et disponibilité comme facteurs déterminant la durée de stimuli verbaux, *Bull. de Psychol.*, 13, 1960, 406-412.

[20] C. E. Noble, The meaning-familiarity relationship, *Psychol. Rev.*, 60, 1953, 89-98.

A developmental study by Hoffmann compared children in grades 1 through 8 and adults on tachistoscopic viewing of (1) groups of consonants, (2) nonsense-syllables (pronounceable), (3) unfamiliar words, and (4) familiar words.[22] The number of letters read correctly increased from condition (1) through (4). The developmental curves differ, as well. The slope is least steep for consonants and increases with the conditions. In other words, groups of consonants are not recognized much better by adults than by first-graders, but the advantage increases as approximation to familiar linguistic units increases.

Miller, Bruner, and Postman compared accuracy of visual recognition for groups of letters having four orders of 'approximation to English.'[23] Accuracy of recognition was greater as approximation to English increased. That frequency as such of either single letters, bigrams, etc., is the critical principle here is unlikely. When the four lists are examined, it is obvious that pronunciability increases with order of approximation to English, and with each stage, more and more 'real' syllables appear. Pronounceable and familiar sub-groups may have played a very important role in increasing accuracy.

Trigram-frequency. Postman and Conger investigated the effect of trigram-frequency (sequences of three letters which appear as parts of words with different frequencies) on recognition-threshold.[24] Two experiments failed to show any relationship between recognition-thresholds and trigram-frequency.

An experiment by Taylor investigated the effect of experimentally produced frequency on trigrams.[25] The trigrams were chosen so as to be approximately equal for summed (individual-letter) frequency. They were selected also for easy pronunciability. An experimental list was presented to the *S*s 15 times. *S* looked at each letter-group for 2 sec. and pronounced it aloud. Afterward, duration-thresholds were obtained for these and for a matched control-list. Frequency of practice of this type did lead to a lowered recognition-threshold. Thus, when a letter-group is viewed and pronounced as a unit, frequency is effective.[26] But mere frequency of sequences embedded in words is not.

Trigrams which form sub-units (that is, pronounceable syllables) of words are another matter. Postman and Rosenzweig compared recognitive thresholds for sets of trigrams which were high and low, respectively, for frequency of occurrence both as actual syllables in English words and as trigrams.[27] The more 'familiar' syllables had lower mean recognitive thresholds. In fact, English words yielded no lower thresholds than familiar syllables.

Meaningfulness. Whether 'meaningfulness' has any role in word-recognition has been the subject of a few experiments. The earliest of these by Peixotto did not use a short-exposure technique, but the results are relevant.[28] She used syllables scaled for association-value from the lists of Glaze,[29] Hull,[30] and Krueger.[31] The *Ss* simply read a list of 30 syllables printed in columns on pages of a booklet. After reading through three pages (the same list repeated in random order) they found syllables they had read mixed with 90 others on the fourth page. They marked the ones they recognized as having seen before. She found little or no relationship between associative value and recognitive score.

---

[21] A. B. Kristofferson, Word recognition, meaningfulness, and familiarity, *Percep. mot. Skills,* 7, 1957, 219-220.

[22] Jakob Hoffmann, Experimentell-psychologische Untersuchungen über Leseleistungen von Schulkindern, *Arch. gesam. Psychol.,* 58, 1927, 325-388.

[23] G. A. Miller, J. S. Bruner, and Leo Postman, Familiarity of letter sequences and tachistoscopic identification, *J. gen. Psychol.,* 50, 1954, 129-139.

[24] Postman and Conger, *op. cit.,* 671-673.

[25] J. A. Taylor, Meaning, frequency, and visual duration threshold, *J. exp. Psychol.,* 55, 1958, 329-334.

[26] See also Solomon and Postman, *op. cit.,* 195-201.

[27] Postman and Rosenzweig, *op. cit.,* 209-226.

A different criterion of meaning was used by Taylor, who had *Ss* study syllables accompanied by a picture of a familiar object which was designed as their 'meaning.'[32] Recognitive thresholds were not lowered for this group as compared with one which had equal frequency of practice with the syllable without the picture (see above). Neither did the degree of associative value of the syllable (as determined by Glaze) have any effect upon performance. She concluded that the variable of meaningfulness, as defined by these procedures, did not influence visual thresholds of duration.

Kristofferson[33] found a significant correlation between meaning (as scaled by Noble[34]) and threshold for word-recognition. In this case, however, meaning covaried with 'familiarity.' It is probable that the latter variable was the effective one, in view of the usual positive results with word-frequency and the negative ones for meaningfulness.

The finding of Postman and Rosenzweig that three-letter English words were not recognized any sooner than familiar syllables (sub-units) of words is consistent with the rejection of 'meaning' as a variable in word recognition.[35] Their words all had a referent (*e.g. bun, pen,* etc.), whereas the syllables did not (*e.g. ing, est,* etc.). They commented that "the failure of the English words to yield lower thresholds than the nonsense-syllables suggests that *S* is no less ready to use syllables as response-units than he is English words of comparable linguistic frequency."

Summary. These experimental results taken together seem to us consistent with our hypothesis that letter-groups which have an invariant spelling-to-sound relationship form functional units for reading. Frequency in one sense would be a necessary condition for the formation of such units, for invariance cannot be discovered without repetition. It would be expected, therefore, that whole words of high frequency and syllables (sub-units of words having a consistent pronunciation) would have high recognitive value, whereas mere sequential frequency would not unless it had a consistent acoustic match. The evidence to be drawn from these experiments is, however, merely supportive; a real test of the hypothesis remains to be made.

---

[28] H. E. Peixotto, The recognitive value of three hundred nonsense syllables, this JOURNAL, 61, 1948, 352-360.

[29] J. A. Glaze, The associative value of nonsense syllables, *J. genet. Psychol.*, 35, 1928, 255-267.

[30] C. L. Hull, The meaningfulness of 320 selected nonsense syllables, this JOURNAL, 45, 1933, 730-734.

[31] W. C. F. Kruger, The relative difficulty of nonsense syllables, *J. exp. Psychol.*, 17, 1934, 145-153.

[32] Taylor, *op. cit.*, 329-334.

[33] Kristofferson, *op. cit.*, 219-220.

[34] Noble, An analysis of meaning, *Psychol. Rev.*, 59, 1952, 421-430.

[35] Postman and Rosenzweig, *op. cit.*, 216.

PRELIMINARY EXPERIMENT

The purpose of the experiment was to test the hypothesis that skilled readers would discriminate visually pseudo-words constructed according to the rules of spelling-to-sound correlation better than words which are not, or are only partially, so constructed. Three lists of pseudo-words were constructed, one with an invariant relation between spelling and sound (*e.g.* NOOSH), one in which all the words were pronounceable, but there were two alternative pronunciations (*e.g.* DRIEND), and one in which the words had low spelling-to-sound correlation (were relatively unpronounceable by rules of English pronunciation, *e.g.* SCRIGW). Word-length varied, but was equated for the three lists. There were 15 words in each list.

The words were presented tachistoscopically, one at a time for 100 m.sec. each. The three lists were mixed in a random order. Items were projected on a screen so that *Ss* could be run in groups. (A second projector and tachistoscope, synchronized with the first, provided a pre- and post-exposure field of a brightness equal to that of the background of the projected word.) The preëxposure field contained two horizontal lines so arranged as to frame the projected word if they were superposed on it. These lines took the place of a fixation-point. The slides containing the words were prepared by typing in bold-face type with a sign-typewriter (typed letters were ⅜ in. high).

The *Ss* were given a 'ready' signal, fixated the lines, and after the word was flashed, tried to write what they saw. They were told to guess when they could. Each word was exposed only once. There were 21 *Ss*, members of a class in the Psychology of Language.

Results were scored right or wrong for the whole word (scoring of parts was tried but discarded as being difficult and no more informative). The mean number correct was about twice as great for the two pronounceable sets of words as for the unpronounceable. A Tukey-test for multiple comparison of means showed that the means for the two pronounceable series were both different from the unpronounceable, but not from each other. If these two had differed, it might have been hypothesized that two alternate pronunciations interfere with one another and reduce discriminability in comparison with a single 'good' pronunciation. This was not the case.

EXPERIMENT I

The hypothesis tested in Experiment I was the same as in the preliminary experiment, but only two sets of words were constructed: one with high spelling-to-sound correlation (referred to as the *P* or pronounceable list) and one with low spelling-to-sound correlation (the *U* or unpronounceable list). Summed letter-frequency, which has not been controlled in the preliminary experiment, was equated for the two lists.

Construction of lists. The 25 words in the pronunceable list consisted of (1) an initial consonant-spelling having a single regular pronunciation, (2) a final consonant-spelling having a single, regular pronunciation, and (3) a vowel-spelling placed between the two consonant-spellings, and having a single regular pronunciation when it follows and is followed by the given initial and final consonant-spellings, respectively. An example would be B/1, E/3, SKS/2. Some of the consonant- and vowel-spellings were single letters, some were 'clusters.' A cluster-type of vowel-spelling may actually contain a consonant-letter as in 'aw.'[36]

The words in the second list—those with low spelling-to-sound correlation—were constructed from the words in the first list. The initial consonant-spelling of the pronounceable words became the final consonant-spelling of the unpronounceable words. Similarly, the final consonant-spelling of the pronounceable words became the initial consonant-spelling of the unpronounceable words. The position of the vowel spellings remained the same.[37] For example: Pronounceable, B E SKS; Unpronounceable, SKS E B. The complete list of words appears in Table I.

TABLE I

CORRECT RESPONSES AND PRONUNCIABILITY-RATINGS FOR PRONOUNCEABLE (*P*) AND UNPROUNCEABLE (*U*) WORDS

	P words	Total correct	Pronun.-rating	*U* words	Total correct	Pronun.-rating
(1)	DINK	118	1.06	NKID	105	5.85
(2)	VUNS	110	1.53	NSUV	90	6.34
(3)	GLOX	117	1.59	XOGL	108	6.12
(4)	SULB	95	1.95	LBUS	115	5.64
(5)	LODS	107	1.28	DSOL	118	4.99
(6)	BESKS	79	2.30	SKSEB	50	6.77
(7)	CLATS	91	1.47	TSACL	102	6.05
(8)	BRELP	91	3.01	LPEBR	35	7.72
(9)	FRAMB	112	2.40	MBAFR	36	7.31
(10)	GRISP	113	1.62	SPIGR	77	5.24
(11)	FUNTS	119	1.62	NTSUF	77	6.10
(12)	TILMS	97	2.18	LMSIT	70	6.44
(13)	SLAND	120	1.71	NDASL	69	6.73
(14)	BLASPS	67	2.88	SPSABL	35	7.28
(15)	SPRILK	38	3.21	LKISPR	14	7.40
(16)	SMAWMP	3	6.20	MPAWSM	20	7.23
(17)	BLORDS	54	2.29	DSORBL	49	6.00
(18)	PREENT	56	2.86	NTEEPR	7	7.47
(19)	DRIGHK	36	6.34	KIGHDR	0	6.72
(20)	KLERFT	37	4.18	FTERKL	33	6.77
(21)	GLURCK	51	2.94	CKURGL	37	6.59
(22)	QUEESK	45	3.05	SKEEQU	35	4.63
(23)	TIRPTH	42	4.50	PTHIRT	31	6.12
(24)	PRILTHS	28	4.63	LTHSIPR	0	8.36
(25)	TRILFTHS	13	5.29	LFTHSITR	0	8.50

The reader may ask, at this point, whether one list is, in fact, pronounceable while the other is not. Actually, the term pronounciability is used here as a convenient means of referring to a correlation of spelling-with-sound, according to the rules of English spelling and pronunciation. One can make an attempt to pronounce 'TSACL' some way or other, but there is no invariant correlation between this spelling and a standard way of pronouncing it in the English language.

---

[36] Spelling must often be 'mapped' into phonemes by clusters or groups of letters, rather than by single letters. Rules for invariance are thus phrased in terms of 'consonant-spelling' and 'vowel-spelling.'

[37] To obtain a pronounceable and an unpronounceable list using the same letters, it was necessary to choose initial consonant-spellings that are not also final ones, and vice versa; or at least one of the consonant-spellings in each pair had to be of this type.

Procedure. The task required of *S* was to identify a word as it was projected on a screen, and to write it down. The procedure was similar to the preliminary experiment in the method of projection and apparatus employed; but it differed in that five presentations of the lists were given, with an increasing exposure-time for each presentation.[38] The 50 words were projected in a random order for 30 m.sec. each. They were then reordered and projected for 50, 100, 150, and 250 m.sec. The words were reordered after each series. The *Ss* were seated about 4 ft. from the screen and were told to write down what they saw and to guess when they could. They were run in small groups to permit good viewing angles for all *Ss*. The height of the letters projected on the screen was 3¾ in., the width about 2½ in. The entire field of projection was 82 in. wide. There were 25 *Ss* (adult, summer-school students); their native language was English.

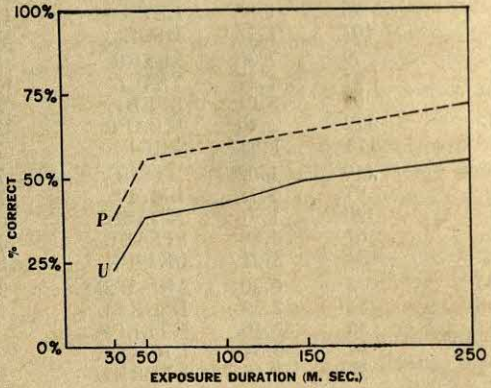


FIG. 1. MEAN PERCENTAGE OF P AND U WORDS CORRECT AT 5 EXPOSURE-DURATIONS

Results. The words recorded by the *Ss* were scored right or wrong, and separate means for *P* and *U* words were obtained for each time-interval. The mean percentage of *P* words correct and *U* words correct is plotted for each of the five successive exposure-durations in Fig. 1. The curve for *P* words is consistently higher than the curve for *U* words. The percentage correct rises sharply from the first exposure to the second, and thereafter increases very little. The two curves are parallel throughout.

The Wilcoxon Matched Pairs Signed Ranks Test[39] was used to compare the total *P* and *U* scores (see Table I). The pairs in each case were composed of the two pseudo-words constructed from the same group of letters, one pronounceable, one not. The difference between the total *P* score and the total *U* score is significant at better than the 1% level (two-tail).

---

[38] This technique was copied from Miller, Bruner, and Postman, *op. cit.*, 129-139.

[39] F. Wilcoxon, *Some Rapid Approximate Statistical Procedures*, 1949.

It was also possible to test the consistency of the trend for *S*s, with Wilcoxon's test. When the scores (number correct) of the *P* and *U* words were compared for the *S*s, the difference between them was significant beyound the 1% level. The difference between them at every time-interval was also significant beyond the 1% level (two-tail test in all cases).

Because of a long established habit of reading from left to right, it might be expected that more errors would occur for the final consonant-spelling than for the initial one. The *S*s' reproductions were scored for initial and final consonant-clusters separately, and this expectation was borne out. But the difference between words of the *P* and *U* lists was still present ($p < 0.01$ by a chi-square test for both initial and final clusters).

Errors can also be expected to increase with length of word. Such was the case. Errors were fewest in the case of four-letter words and increased with length. The difference between *P* and *U* words occurred for all lengths of word but was smallest for four-letter words and greatest for five-letter words. The interaction could not be measured, since the number of cases for the different lengths of word was not equal, but the difference is quite obvious. A short word of four letters is so seldom missed that the difference between *P* and *U* can have only a small effect. On the other hand, a word of eight letters is so difficult that not even pronunciability will permit perfect discrimination to occur with great frequency.

The errors were studied and classified also for the kinds of changes made. 'Real' word-errors occurred infrequently, probably because the *S*s were told in the instructions that they would be shown 'nonsense' words. More real words were given for *P* combinations than for *U* combinations (27 in all, as against 15).

Errors which changed the projected letter-group in the direction of a more pronounceable one were frequent. This was accomplished in some cases by adding a vowel (*e.g.* NIKID for NKID), by omitting a consonant (*e.g.* SKEB for SKSEB), or by changing a consonant-cluster (*e.g.* BLUS for LBUS). An error of this type was often repeated on later trials.

Errors of omission (writing nothing) were more frequent for the unpronounceable words, but such errors did not occur often. They occurred most often for the longer words.

Discussion. The results of this experiment demonstrate that a letter-group with a high spelling-to-sound correlation is reproduced more accurately than an equivalent letter-group with a low spelling-to-sound correlation.

This result cannot be caused by a difference in the familiarity of the letters taken alone, or even the vowel- and consonant-clusters taken alone, for the same clusters were used in the two lists. It must be due to the existence of higher-order graphic units: the letter-combinations of English writing that function as relatively stable units in grapheme-phenome correspondences.

Practically, this result suggests strongly that the proper unit for analyzing the process of reading (and writing) is not the alphabetical letter but the spelling pattern which has an invariant relationship with a phonemic pattern. This may be of great importance for children's learning to read and write.

Among the theoretical implications for psychology, perhaps the most general is this, that while reading is based on the discrimination and identification of visual forms such as letters, it becomes, in the skilled reader, a process of perceiving 'super-forms,' and that these tend to be constituted (organized) by their relation to auditory-vocal temporal patterns. Insofar as frequency has a role in the constitution of these units, it is the frequency of grapheme-phoneme coincidence which is crucial, not frequency of exposure to the seen or uttered units alone. The reading of words is thus inseparable from the hearing of words. Since the hearing of words is also inseparable from the speaking of words, reading must be conceived, however, as part of a circular response-process, not simply as a stimulus-response process. The stimuli are complex, and the eventually effective stimuli are response-produced, in the sense that the visual stimuli which come to function as units are structured partly by feedback from response. Verbal perception, unlike certain other kinds of perception, will have to be explained in terms of a still emerging theory of self-stimulation or feedback.

EXPERIMENT II

The above interpretation of the results in Experiment I assumes that the pseudo-words having high spelling-to-sound correspondences are actually perceived (read) more quickly and accurately than those with low correspondence. Proponents of theories emphasizing *emitted* frequency or 'response-bias' will be tempted to object that the difference between the two sets of words was not due to a difference in perceptibility, but instead due to the *S*'s tendency to emit, whatever the stimulus presented, the responses which were in the past made most frequently.[40]

---

[40] See, for example, Israel Goldiamond and W. F. Hawkins, Vexierversuch: The log relationship between word-frequency and recognition obtained in the absence of

The 'spew hypothesis' of Underwod and Schultz[41] states the operation of response-bias as follows, "the order of emission of verbal units is directly related to frequency of experience with those units." Are the pronounceable words recorded correctly more often by the *S*s only because they are nearer the threshold of availability as responses, or are they actually perceived better?

If emitted frequency, in the past, is responsible for the words recorded, one might expect that *errors* in the previous experiments would include many which changed spelling without changing pronunciation. An example would be writing KLERPHT for KLERFT or GLURK for GLURCK. Some errors of this type were made, but fewer than seem consistent with a serious consideration of the hypothesis. Only 16 errors (out of a total of 7250 opportunities to respond) could be classified as identical in pronunciation with the pseudo-word exposed on the screen. But there are so many possible causes of error that one has no expected frequency with which to compare this number. A more direct test therefore seemed advisable.

If the *S* is not required to *emit* a word-response at all, but is instead asked to recognize or match one word with another, we can be fairly sure we are testing what he has perceived, not merely the response which is most available and easy to emit. He does not have to produce the response at all, in this case. One could use a judgment of same or different, or a match from a list including the word shown and some incorrect versions.

Our second experiment was designed to test the hypothesis that an *S* would *recognize* the word shown, from a multiple-choice list including this word, more accurately when the word shown had a high spelling-to-sound correlation.

Construction of lists. The pseudo-words exposed for recognition were 20 from the two lists (pronounceable and unpronounceable) used in Experiment I. The 'misleads' for the multiple-choice series for each one posed a problem. Artificial construction of these by made-up rules would yield unknown and probably unequal difficulty for different words. Also, the S might discover some of the rules before finishing and thereafter choose on a problem-solving basis, rather than perceptual matching. The series was obtained empirically, therefore, by including for each word the three most frequent errors recorded in the previous experiment, whatever they were. For example: for GLURCK the following words were used: GLURK, GLUCK, GLURGK, and GLURCK; for CKURGL the following were used: CKUGL, CKURL, CKUGEL, and CKURGL. Since one position in the choice-list might be more favorable than

---

stimulus words, *J. exp. Psychol.*, 56, 1958, 457-463; C. R. Brown and Herbert Rubenstein, Test of response bias explanation of word-frequency effect, *Science,* 133, 1961, 280-281.

[41] B. J. Underwood and R. W. Schulz, *Meaningfulness and Verbal Learning,* 1960, 86.

another, four lists were so constructed for each word that each choice appeared once in each of the four positions.

Practice-effects were equated by making two random orders of the words to be presented tachistoscopically, the second one so constructed that the unpronounceable version of a pronounceable combination now appeared in the position the pronounceable one had occupied in the first order (and vice versa). There were therefore 8 sub-groups constituted by the two orders, each with four versions of the multiple-choice series.

Procedure. *S* was seated 5 to 7 ft. from a screen on which the words were projected, as before. The width of the total field projected was 31 in. The width of the word on this field varied from 12 to 24 in., depending on number of letters. The letters, as projected, were 3½ in. high. *Ss* were run individually. The instructions read to *S* were as follows:

> On the screen before you, between the two black lines, a nonsense-word will appear for only a fraction of a second. Immediately afterward, a list of four nonsense words will appear, *one* of which will be the word just exposed.
>
> Choose the one you think is the same as the word you have just seen and place a number in the blank corresponding to its order on the screen. Number 1 is at the top and Number 4 at the bottom. A 'ready' signal will be given before the first exposure. When the 'ready' signal is given, please be sure to fixate between the two lines so as not to miss the word.
>
> Do not try to make inferences about the alternatives provided in the choice-list, because they were obtained experimentally. Just look at the word flashed, and pick the one in the list which looks most like it.

TABLE II

NUMBER OF CORRECT CHOICES OF *P* AND *U* WORDS SEPARATED FOR ORDER AND MULTIPLE-CHOICE MATCHING LIST

Order	Words	List 1	List 2	List 3	List 4	Mean
A	*P*	121*	171†	105*	106‡	14.37
	U	92*	130†	83*	91‡	11.31
B	*P*	96§	91§	84§	103‡	14.96
	U	63§	74§	68§	88‡	11.72
Mean	*P*	15.50	14.56	13.50	14.93	14.62
Mean	*U*	11.07	11.33	10.79	12.79	11.48

* *N*=8; † *N*=12; ‡ N=7; § *N*=6.

Ten practice-words were given first to accustom *S* to the timing of the exposure. A pseudo-word was exposed tachistoscopically for 30 m.sec. and was followed after ½ sec. by projection of the appropriate choice-list. This list was projected for a duration of 8 sec. The duration was long enough to permit *S* to read the words and record one, but not to study the alternatives in detail.

The 40 experimental words followed, in the order assigned to the sub-group. The entire experiment lasted about 18 min. There were 60 *Ss*, all university students. Each sub-group contained at least 6 *Ss*. An *S* took part in only one sub-group.

Results. The results were first scored separately for each sub-group, to see whether trends changed depending on order in the whole series or position in the choice-list. Table II contains the results by sub-groups. The mean number correct with a given order can be read across the table. The mean number correct for a given order in the choice-list can be read downward. For every sub-group (Order A combined, Order B combined, and Choice-Lists 1, 2, 3, and 4) the pronounceable words were matched correctly more often than the unpronounceable words. It seemed justifiable, therefore, to combine all the results. The over-all means were 14.62 correct for pronounceable words and 11.48 correct for unpronounceable words.

Since a quarter of the choices would be expected to be correct by chance, the number perceived correctly beyond chance was 9.62 for *P* words and 6.48 for *U* words. This is a gain of 48% for *P* words over *U* words.

When the words (pronounceable and unpronounceable) were matched by pairs, 15 of the 20 pronounceable words were perceived correctly more often than their unpronounceable mates. In four cases, the unpronounceable version was perceived correctly more often, and in one case there was no difference. By the Wilcoxon Matched-Pairs Signed-Rank Test (two-tailed) the difference between the lists was significant at $p < 0.002$. The differences were also significant ($P < 0.01$) when tested by *S*s.

The results of this experiment, then, corroborate those of the previous one and provide further positive evidence that pseudo-words constructed according to rules of invariant spelling-to-sound correlation of the English language are perceived more accurately under conditions of tachistoscopic viewing than their matches with no invariant spelling-to-sound prediction. It seems reasonable to conclude that the *perception* of the stimulus-word is affected by this correspondence; that we are not observing simply a 'response-bias' to emit pronounceable combinations.

When responses must be emitted, however (as in Experiment I), there might conceivably be a combined effect of the perceptual bias toward grasping the higher-order invariant and a response-bias toward pronunciability. In that case, the difference between the two word-lists in Experiment I should be considerably greater than that in Experiment II. It is impossible to make a direct comparison here, since a longer list (25 pairs) was used in Experiment I, as well as a different method (reproduction vs. recognition). The advantage of the pronounceable words over the unpronounceable words at the first exposure (30 m. sec.) was 65% in Experiment I, as against 48% in Experiment II. It is possible, therefore, that both effects are operative in Experiment I, resulting in greater difference in the results.

The question now to be answered is how the rules of invariance for grapheme-phoneme correspondence are learned. To answer this question it may be feasible to 'build in' such units experimentally. If this can be done, the learning process presumably could be studied. Meanwhile, a simple associational model does not seem the best to entertain, for the rules of correspondence are conditional (whether a cluster is at the beginning or the end or the middle, etc.). A trigram such as 'ign' functions as a predictable cluster for pronunciation at the end of a word (*e.g.* sign, malign) but the prediction will be different in another context (*e.g.* significant). A model more like a paradigm for concept-formation, with conditional isolation from context, seems more promising. Hopefully, a workable model can be formulated. Developmental experiments which may reveal the presence or absence of functional units of the kind posited at different age-levels are now in progress, and should provide some clues.

CONTROL EXPERIMENTS ON PRONUNCIABILITY

Rating of pronunciability. Are letter-combinations which have, according to the rules, an invariant grapheme-phoneme correspondence, actually more 'pronounceable'? The term 'pronunciability' has been used in this paper as if such a statement could be made. But the matter is not merely one of convenience of reference, since Underwood and Schulz found that pronunciability, as defined by their scaling procedure, was a very important variable in verbal learning.[42] Ratings were therefore obtained of the pronunciability of the words in the *P* and *U* lists.

The rating method devised by Underwood and Shulz was copied verbatim except for references to word-length and 'common' words (some of theirs were common words, while all of ours were pseudo-words). The two sets of words were mixed and presented to a group of 165 *Ss* (students in an introductory class in psychology) for rating on a 9-point scale. A score of 1 equals high and 9 equals low. Order was reversed for half the *Ss* in case scale changes would occur with practice.

The results showed a clear difference in pronunciability between the two lists. Mean ratings were computed for each of the 50 pseudo-words (see Table I). The over-all mean rating of the pronounceable words was 2.88 ($SD = 1.48$); that of the unpronounceable words was 6.57 ($SD = 0.94$). For every one of the matched pairs the pronounceable combination received a lower mean rating (*i.e.* more pronounceable) than the unpronounceable combination. In all cases the mean for a pronounceable combination was lower than the over-all mean for the unpronounceable ones; similarly, the rating of each unpronounceable combination was higher than the over-all mean of the pronounceable combinations. There was considerable inter-individual variability in rating, probably related to differences in background of language, but the results fully justify the characterization of one list as more pronounceable than the other.

Invariance of pronunciation. A second question may be asked regarding the actual pronunciation of the pseudo-words. The rules of grapheme-phoneme correspondence imply that there should be a relatively invariant pronunciation of the *P* words, but a variety of pronunciations for the *U* words, which do not follow the rules. Is this indeed a fact? To answer the question, 16 Ss were asked to read the

---

[42] Underwood and Schulz, *op. cit.*, 163-199.

50 words aloud. They first read quickly aloud through the two lists, one and then the other. The total time was noted for each list. Then they read the words, mixed in a random order, with a 5-sec. pause between words. The words were exposed, one at a time, on cards. A tape recording was made for each *S*. Order was balanced among the *S*s.

The first reading served both as practice and as a time-check. If readability under conditions of tachistoscopic exposure is related to reading skill under normal conditions, the total reading time should differ for the two lists. The mean reading time was 38.8 sec. for the *P* list and 57.5 sec. for the *U* list, thus confirming this expectation. The trend was consistent for individual *S*s.

The tape recordings (all 50 words for each of the 16 *S*s) were given to two linguists for analysis of the pronunciations. They worked individually. The method used on their first transcriptions was partly phonetic and partly phonemic. The sequences of letters in the *U* list, and perhaps occasionally in the *P* list, were such that the *S*s sometimes brought in phonemes from other languages to make them easier to pronounce. In general, however, the *S*s included extra sounds, so as to push the sequence toward a possible English one. On their second analysis, the linguists tried to put all the phonetic transcriptions in terms of the phonemic system of English (that is, in terms of significant contrasts). When the two linguists had disagreed on a transcription, the word was played several times and usually an agreement resulted. There was finally 98% agreement between them.

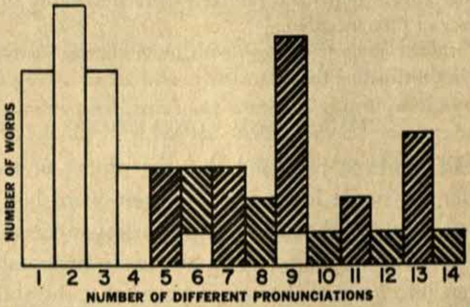


FIG. 2. DISTRIBUTION OF FREQUENCIES OF *Different* PRONUNCIATIONS FOR *P* AND *U* WORDS
(Cross-hatching indicates *U* words)

Fig. 2 shows the distributions of different pronunciations for the *P* list and the *U* list. There is very little overlapping, though two *P* words (DRIGHK and SMAWMP) were quite inconsistently pronounced and fell within the *U* distribution. The mean of different pronunciations was 2.68 for the *P* words and 9.12 for the *U* words. When *P* words were compared with their *U* counterparts, all *P* words had fewer different pronunciations except one (DRIGHK) which had a number equal to its counterpart.

The words in the *U* list were invariably pronounced as bisyllabic words, whereas the words in the *P* list were almost always pronounced as monosyllables.

It can be concluded that pseudo-words constructed according to rules of grapheme-phoneme correspondence are pronounced more consistently when read aloud than those which are not.

Relation between scaled pronunciability and invariance of pronunciation. Whether pronunciability, as measured by a rating scale, is related to invariance of pronunciation (relative to rules of grapheme-phoneme correspondence) can be determined by comparing the results of the two experiments above. The results are independent, since none of the *S*s took part in both experiments.

Both sets of scores (ratings and variability of pronunciation) for the 50 words were transformed into standard-scores and correlated by means of a Pearson *r*. A positive correlation of 0.85 was found (not corrected for attenuation). The magnitude of this correlation is very interesting, for it suggests that the rules of grapheme-phoneme correspondence, which exist objectively in the language, are reflected psychologically in individual ratings of pronunciability of words and also in group conformity of pronunciation.

These two experiments seem to us to provide quite impressive empirical validation of the rules used in constructing the lists.

SUMMARY AND CONCLUSIONS

The hypothesis has been proposed that the proper unit of the reading process is neither the single letter nor the whole word but a higher-order invariant derived from grapheme-phoneme correspondences. Rules can be drawn up for predicting spelling from sound if temporal patterning of 'clusters' is taken into account. It was proposed that the skilled reader has learned to perceive as units the letter-patterns conforming to these rules and that such patterns have, therefore, an advantage in reading.

Several experiments were performed to test this hypothesis. Pseudo-words conforming to rules of spelling-to-sound correlation were compared with ones of low correlation under conditions of tachistoscopic viewing.

In the first experiment, it was found that the *S*s reproduce pseudo-words of high correlation more accurately than ones matched for summed letter-frequency, but of low correlation.

In a second experiment, comparable results were obtained when the *S*s were not asked to reproduce a word, but performed perceptual matching of the word exposed tachistoscopically with a member of a multiple-choice list. It was concluded that the perceptual process has been facilitated in skilled readers for units discovered during long exposures to the grapheme-phoneme correspondences of the English language.

MEDIATED RESPONSES TO SIZE AND BRIGHTNESS AS A FUNCTION OF AGE

By TRACY S. KENDLER, Barnard College, HOWARD H. KENDLER,
New York University, and BEULAH LEARNARD,
Columbia University

For those who investigate higher mental processes within the domain of S–R psychology, the concept of the mediating response is of singular importance.[1] It is usually treated as a response which intercedes between the external stimulus and the overt response to provide stimulation that influences the course of overt behavior. Since this response is, ordinarily, more often implicit than explicit, its investigation poses problems. One method of dealing with such problems is to develop techniques that make the mediating responses overt, as in the measurement of movements of the tongue during the solution of problems in arithmetic.[2] Another method, exemplified by the present study, is to infer the process from overt behavior as manifested under a prescribed set of conditions. In this latter method, the mediated response has the status of an hypothetical construct which may or may not be ultimately measurable, but which justifies its use by its theoretical and heuristic function.[3]

The general technique used in this investigation is based on procedures developed by Buss, and Kendler and D'Amato.[4] It consists essentially of studying mediation by means of the transfer demonstrated from an initial to a subsequent discrimination. The initial discrimination presents stimuli that differ simultaneously on two dimensions, but only one of the dimensions is relevant. After criterion is reached, another

---

* Received for publication April 10, 1961. This experiment is part of a program of research supported by a grant from the National Science Foundation. The valuable coöperation of Public Schools 93 and 105, the Gardens Nursery School, the Riverside Nursery School, and the Open Door Child Care Center, all of New York City, and the Hawthorne Public School of Massapequa, New York, is gratefully acknowledged.

[1] C. N. Cofer and J. P. Foley, Jr., Mediated generalization and the interpretation of verbal behavior: I. Prolegomena, *Psychol. Rev.*, 49, 1942, 513-540; C. L. Hull, Knowledge and purpose of habit mechanisms, *Psychol. Rev.*, 37, 1930, 511-525; N. E. Miller and John Dollard, *Social Learning and Imitation*, 1941, 69-90; C. E. Osgood, A behavioristic analysis of perception and language in cognitive phenomena, in *Contemporary Approaches to Cognition*, 1957, 75-118.

[2] Study by Sokolov, reported by A. Mintz, Further developments in psychology in the U.S.S.R., *Ann. Rev. Psychol.*, 10, 1959, 464.

[3] H. H. Kendler, Problems of problem-solving research, in *Current Trends in Psychological Theory: A Bicentennial Program*, 1961.

[4] A. H. Buss, Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations, *J. exp. Psychol.*, 45, 1953, 75-81; H. H. Kendler and M. F. D'Amato, A comparison of reversal shifts and nonreversal shifts in human concept formation behavior, *ibid.*, 49, 1955, 165-174.

discrimination is presented that utilizes the same or similar stimuli but requires a shift in response. One type of shift, called a reversal-shift, requires *S* to continue to respond to the previously relevant dimension but in an opposite way. In another type of shift, called a non-reversal shift, *S* is required to respond to the previously irrelevant dimension. For example, if *S* is initially trained on stimuli that differ in brightness (black vs. white) and size (large vs. small) by being rewarded for responses to black, regardless of size, a reversal-shift would consist of learning to respond to white and a non-reversal shift would consist of learning to respond to small. Comparison between these two types of shifts is of particular interest because theories based on single-unit and on mediated S–R connections yield opposed predictions about their relative efficiency. A single-unit theory that assumes a direct association between the external stimulus and the overt response would predict reversal-shift to be more difficult than non-reversal shift. This is because reversal-shift requires the replacement of a response that has previously been consistently reinforced with a response that has previously been consistently extinguished. In a non-reversal shift, previous training has reinforced responses to the newly positive and newly negative stimuli equally often. Strengthening one of these associations does not require as much extinction of its competitor as in a reversal-shift and should therefore be acquired more easily. Kelleher confirmed the prediction that, for rats, reversal-shift was more difficult than non-reversal shift.[5]

A theory that includes a mediating link (or links) between the external stimulus and the overt response leads to a different prediction. The mediating link is conceived of as a perceptual or verbal response, often covert, to the relevant dimension which produces cues that determine the overt response. In a reversal-shift, the initial dimension maintains its relevance; hence, so does the mediated response. Only the overt response needs to be changed, and since the experimental situation provides only one alternative overt response, this change presents no great difficulty. In a non-reversal shift, the previously acquired mediation is no longer relevant, consequently both the mediating and the overt response must be replaced, making this a more difficult task than a reversal-shift. Experiments by Buss, Kendler and D'Amato, and Harrow and Friedman, using a more complex variation of the technique of reversal and non-reversal shift with college students as *S*s, confirmed the prediction of the mediational analysis.[6] Unlike rats, college students acquire a reversal-shift more easily than a non-reversal shift.

This discontinuity between rats and adult humans led to two investigations with young children to determine whether their behavior, in this type of situation, was more consistent with the single-unit or mediational formulation. The results suggested that children between 3 and 4 yr. of age respond predominantly in the single-unit manner,[7] and that children between 5 and 7 yr. of age divide about evenly with half mediating and half not.[8] These data imply a developmental process in which very young children's behavior is governed by a relatively primitive single-unit *S–R* process, with increasing maturity leading to increases in the proportion of children who mediate. The purpose of the present investigation was to provide a direct test of this implication. Previous procedures were modified to allow each *S* to choose whether they would behave mediationally or not. This was done by presenting a shift that could be accomplished on either a reversal or non-reversal basis. Since the choice was left with the *S*s, each one could be identified as a mediator or non-mediator, depending on whether he chose a reversal or a non-reversal shift. Buss used such a procedure on college students and found that 72% chose a reversal-shift.[9] The present study applied this method to children at five chronological age levels; namely 3, 4, 6, 8 and 10 yr. of age, to determine whether there is an increase in the proportion of children who mediate and, if so, to study the course of this increase.

---

[5] R. T. Kelleher, Discrimination learning as a function of reversal and nonreversal shifts, *J. exp. Psychol.,* 51, 1956, 379-384.

[6] Buss, *op. cit.,* 75-81; Kendler and D'Amato, *op. cit.,* 165-174; Martin Harrow and G. B. Friedman, Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled, *J. exp. Psychol.,* 55, 1958, 592-597.

[7] T. S. Kendler, H. H. Kendler and D. Wells, Reversal and nonreversal shifts in nursery school children, *J. comp. physiol. Psychol.,* 53, 1960, 83-87.

TABLE I

NUMBER, AGE AND ESTIMATED *IQ* OF *S*s IN EACH GROUP

Age in months		*N*	Mean estimated *IQ*
mean	range		
43.3	37–47	24	120.2
54.3	49–59	32	112.0
77.3	73–82	32	112.2
101.7	96–109	32	113.8
127.0	121–134	32	113.6

METHOD

Subjects. The *S*s were 171 children of both sexes drawn from schools in the metropolitan area of New York, ranging in age from 3 to 10 yr. Of these, 19 were eliminated; 8 because their estimated *IQ* was below 80, 8 because of the rough matching in *IQ*, and 3 because they did not learn the initial discrimination within 130 trials. The 3 non-learners were all in the 3- and 4-yr. old groups.

The 152 *S*s who met the requirements of the experiment were distributed as shown in Table I. The *S*s were run in two replications, each with a different *E*.[10] The replications were identical with reference to procedure and experimental conditions, similar with reference to population sampled, and, except for the youngest group, equal in number. In the 3-yr. group, one replication consisted of 16 *S*s and the other of 8.

---

[8] T. S. Kendler and H. H. Kendler, Reversal and nonreversal shifts in kindergarten children, *J. exp. Psychol.*, 58, 1959, 56-60.

[9] Buss, Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated, *J. exp. Psychol.*, 52, 1956, 162-166.

[10] Beulah Learnard was *E* in one replication and Helen S. Pollack in the other.

Apparatus. The discriminative apparatus used was 17 in. high, 10 in. wide and 4 in. deep.[11] Paired windows, manipulanda and slots for delivery of the reward were arranged symmetrically on the frontal surface. The windows, through which the discriminanda were exposed, were 3 × 3 in. The manipulanda were round steel-rods, ⅜ in. in diameter and 4¼ in. in exposed length, long enough slightly to overlap the stimulus-cards. *S* manipulated the rods by pushing them toward the discriminanda, an operation accompanied by a loud mechanical click. The base of each rod was attached to a concealed arm of an electronically controlled marble-magazine, which when activated delivers marbles to the openings beneath the manipulanda.

The discriminanda consisted of four mid-gray pasteboard-cards on which squares that differ in size and brightness were mounted. Two of the squares were black and two were white. Two were small (1 sq. in.) and two were large (3 sq. in.). The four discriminanda were so paired as to vary simultaneously in two dimensions. Thus small black (SB) was always paired with large white (LW) and large black (LB) with small white (SW). The cards were inserted by hand into a grooved container behind the windows. A red plastic screen could be slid into place to cover the windows and conceal *E*'s manipulations from *S*.

Procedure. The *S*s were run individually in a room in which *E* and *S* were alone. The procedure was begun with the administration of selected sub-tests from the Stanford-Binet and the Wechsler Intelligence Scale for Children (*WISC*) in order to provide a quick, albeit rough, basis for equating the *IQ*s of the various age-groups in the experiment. First the vocabulary sub-test of the WISC was given. If a child attained the minimal score (equivalent to an *MA* of 58 mo.) or more, that ended the *IQ*-test. Children who did not attain this level were presented with the following items from the Stanford-Binet test: picture vocabulary, comprehension- and definition-items for years II, II-6, III, III-6, IV, IV-6, and V until a basal level was established. In all cases, tests were administered and scored in accordance with the scale from which they were drawn.

Following the *IQ*-test, *S* was seated comfortably in front of the apparatus and *E* read the following instructions.

> This is a game in which you will try to win as many marbles as you can. Here are some marbles for you to start with. [*E* handed five marbles to *S*.] Put them in the holes in this board. [In front of *S* was a grooved board to hold the marbles he won.]
>
> Now (name of child), [*E* dropped screen] look at these two pictures. The way we play this game is that there will always be one picture in each of these little windows. One picture is a winning picture and the other picture is a losing picture. This is how you find out which picture is a winner. Look at these two sticks. Each stick points to one of the pictures like this. When we start the game you will press the stick that points to the picture that you think is the winner. If you are right, a marble will drop out of one of these little holes and you may put it with the rest of the marbles you have. If you are wrong, no marble will drop out. Instead, you will give one of your marbles back to me. Each time I drop this screen, you will have *one* turn to press. On each of these turns, one of the pictures is a winner and the other picture is a loser. Sometimes the winner will be in this window and sometimes the winner will be in that window. If you try, you can win a marble

---

[11] The apparatus was built to specification by the Scientific Prototype Mfg. Co., Brooklyn, New York.

every time you choose. Remember, each time I drop this screen you will have only one turn to choose.

Ready? Then we'll start. Now press one of the sticks and see if it points to the winning picture. [After the first correct choice, *E* said]: That was the winning picture. Every time you win a marble, pick it up and put it with the rest of your marbles. [*E* pointed to a grooved marble board. At this point, *E* explained that at the end of the game *S* would be allowed to keep one of the marbles but the rest had to be returned so that *E* could play the game with the rest of the children. After the first incorrect choice, *E* said]: That was the losing picture. So you'll have to give back one of your marbles. [After every tenth trial, *E* said]: Don't forget to look at the pictures. Remember, if you try, you can win a marble every time you choose.

Three series of trials were presented with no noticeable break in the procedure between them. Series I provided training in an initial discrimination; Series II, training in a second discrimination; and Series III, the test-trials.

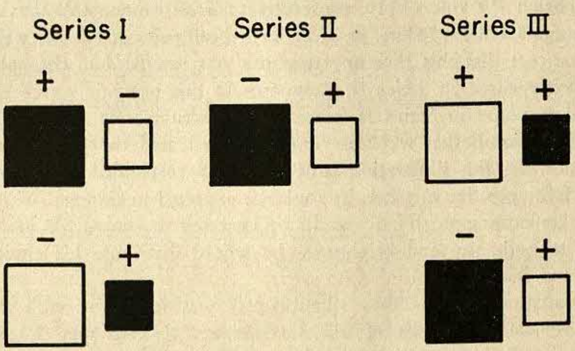


FIG. 1. ILLUSTRATION OF ONE OF THE ARRANGEMENTS OF STIMULI AND REINFORCEMENT USED IN THE EXPERIMENT

In Series I, both pairs of stimuli were presented alternately in a prearranged irregular sequence so designed that (1) each pair appeared equally often but no more than twice in succession; and (2) the correct stimulus appeared equally often on the right and the left but no more than twice in succession in either position. The actual sequence presented to each *S* depended, however, to some extent on his performance. Because children, especially the younger ones, are prone to form position-habits early in learning that can considerably prolong the acquisition of a discrimination, each pair was presented in the same position as many times as it was necessary before the child chose the correct stimulus. After a correct response, the next pair in the prearranged sequence described above was presented. This procedure reduced the intermittent reinforcement that could otherwise sustain a position-habit for a long time.

During this series, a response to one member of each pair was always correct and, consequently, rewarded with a marble. If the response was to the negative member of the pair, *S* was obliged to return a marble to *E*. The positive stimuli shared one aspect of the stimulus and differed on the other. Fig. 1 presents an illustration of one of the arrangements of the stimuli and reinforcements presented in each series, which will be used as an example throughout. In the illustration both positive members are black and size is irrelevant. *S* was run in this series until a criterion of 9 out of 10 consecutive correct repsonses was reached. Each aspect of the stimulus *i.e.* B (black), W (white), S (small), and L (large) was the positive stimulus for one quarter of the *S*s, as arranged by random assignment.

After criterion was reached in Series I, Series II was presented. During this series only one pair of stimuli was presented. For half of the *S*s it was one of the pairs, for the other half it was the remaining pair. For all *S*s the previously negative stimulus became positive and the previously positive stimulus became negative. In this series, the stimuli differed on two dimensions and there was no second pair to make one dimension more relevant than the other. In the illustration provided, the positive stimulus is both small and white. *S* could learn by responding to the size, to the brightness, or to both. The third series was designed to ascertain the basis on which *S* learned Series II.

After criterion (9 out of 10 consecutive correct responses) was reached on Series II, Series III began. Here, as in Series I, both pairs of stimuli were presented alternately, except that this time any response was rewarded to the pair of stimuli that had not appeared in Series II. Responses to this test-pair served to determine the basis of response in Series II, since these discriminanda separated the critical aspects of the stimuli that were previously merged, and the procedure required a choice of one. In the illustration provided, if *S* responded in Series II to the whiteness of the positive member, he could be expected to choose LW in Series III. If he were responding to size, he could be expected to choose SB. If he had been responding to both size and brightness, he would distribute his choices between the members of the pair.

The presentation of the other stimulus-pair was interlaced with the test-pair. It retained the pattern of reward and was designed to help keep *S* responding on the same basis as in Series II. Each pair was presented 10 times in the same alternating sequence as in Series I.

At the completion of Series III, the *S*s were shown the pair from Series II once again and asked some questions designed to provide information about the verbal processes associated with this task. The first question was, "Which one was the winner?" If the child made no relevant verbal response, or if he merely pointed, he was asked, "What does it look like?" If there was still no relevant verbal comment, he was asked, "How do you know?"

After the end of the session, *S* was pledged to keep secret the nature of the 'game' until everyone in his school had had a turn.

Classification. *S*'s choices on the test-pair during Series III provided the basis for classification into one of the following three categories: reversal, non-reversal and inconsistent.

(1) Reversal. If, on the test-pair, *S* made eight or more responses to the stimulus that had been incorrect in Series I, he was classified as a reverser. For example, if, as in Fig. 1, B was positive in Series I and *S* chose LW consistently in the test-series, it would indicate that Series II had been learned by shifting from B to W, or in other words, by making a reversal-shift.

(2) Non-reversal. If, on the test-pair, *S* made eight or more responses to the stimulus that had been irrelevant in Series I, he was classified as a non-reverser. In

our example, this *S* would choose the SB in the test-series, thereby indicating that although Series I was acquired by responding to brightness, Series II had been acquired by responding primarily to the size of the positive stimulus. This learning is based on a shift to the previously irrelevant dimension and has in previous research been called a non-reversal shift.

(3) Inconsistent. Another possibility is that *S*'s choices in Series III may be inconsistent, *i.e.* neither stimulus of the test-pair may be chosen more than seven times. This behavior would indicate that by the end of Series II, *S* was responding equally, or almost so, to the size and brightness of the positive stimulus.

RESULTS AND DISCUSSION

The main concern of this research was with the changes in the proportion of children who used a mediating response as a function of chronological age. Whether a child mediated was inferred from choice-behavior in Series III, and so the analysis of results begins there. The experimental design included two replications with two different *E*s in order to assess the stability of the results. Therefore, a preliminary series of χ^2-tests were made comparing the results of the two replications for each age-level to determine whether the replications required separate treatment or whether they could be treated as though drawn from the same population. None of the χ^2-tests approached statistical significance at the 5% level, consequently the two replications were combined for all the ensuing analyses.

Fig. 2 presents the major results of the experiment, *i.e.* the percentage of children in each category of choice plotted as a function of age. To develop fully the implications of these results, each category will be given separate consideration.

For reasons previously presented, reversers are considered to have made mediated responses. These are the children who, when faced with a shift, continue to respond to the same dimension in a reverse manner, *e.g.* shift from black to white. The expectation about this category, based on previous results and theoretical considerations, was that it would lie below 50% between the ages of three and four,[12] rise to about 50% between five and seven,[13] and then continue to increase with increasing age until some relatively high asymptote is reached.[14] The results were that 37.5% of the 3-yr. olds mediated, and this proportion increased to 50% at 4 yr., remained level at 6, and then rose gradually to 62.5% at age 10 yr. A test of the significance of differences between 3- and 10-yr. olds, based on an estimate of the 'true' proportion drawn from the total group, yielded a one-tail $p = 0.03$. Although the curve appears to be approaching a rather low asymptote, these results confirm the general developmental implications of previous studies and also show an almost surprising measure of agreement where specific comparisons are possible.

---

[12] Kendler, Kendler, and Wells, *op. cit.*, 83-87.

[13] Kendler and Kendler, *op. cit.*, 56-60.

[14] Buss, *op. cit.*, 1956, 162-166.

On the basis of the following brief rationale, derived from Spence's single-unit theory of discrimination-learning,[15] the non-mediators were

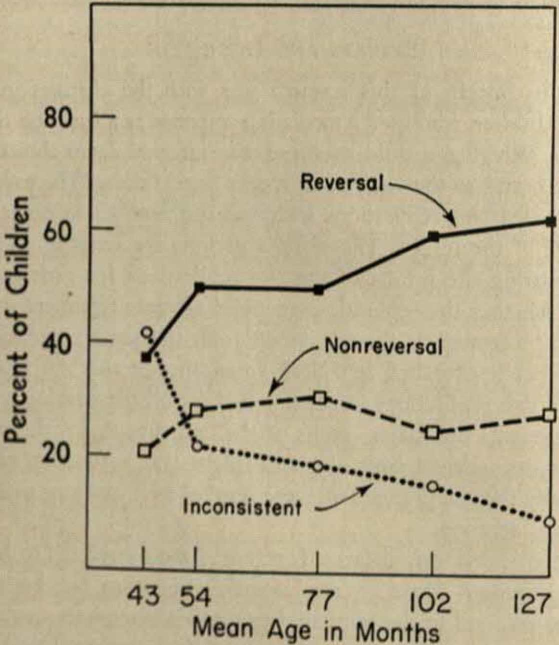


Fig. 2. Percentage of Children at Each Age-Level Responding in Each Category

expected to be distributed between the non-reversal and inconsistent categories. In this framework, each discriminandum is viewed as a complex consisting of two stimulus-components, each of which can acquire its own excitatory tendency as a function of the pattern of reinforcement. Thus at the end of Series I, the excitatory tendency of the positive component of the relevant dimension (*e.g.* B, as in Fig. 1) should be relatively high, since responses to it have always been reinforced. The excitatory tendency of the negative component of the relevant dimension (W) should be relatively low, since responses to it have never been reinforced.[16] The excitatory tendencies of the two components of the irrelevant dimension (L and S), which have been reinforced roughly half of the time, should be of a magnitude somewhere between those of B and W. At the outset of Series II, the child should continue to respond to the stimulus-compound that was previously correct (LB) until the tendencies previously built up become sufficiently weakened through non-reinforcement to permit responses to the other stimulus-compound (SW), which are immediately reinforced. Each component of the newly positive stimulus-compound should have different excitatory strengths as a result of the training in Series I. Thus, the response to W should be weaker than the response to S. If a non-mediating child reaches criterion during the time this differential is maintained and is then required, as in Series III, to choose between W and S, he should choose S predominantly. This response would place him in the non-reversing category.

---

[15] K. W. Spence, The nature of discrimination learning in animals, *Psychol. Rev.*, 43, 1936, 427-449; The nature of response in discrimination learning, *ibid.*, 59, 1952, 89-93.

But the analysis of single-unit behavior is not yet complete. If, as is assumed here, there is no selective process operating in non-mediators, then all stimuli impinging on the child at the time of reinforced response should acquire an increase in excitatory strength for that response. Consequently responses to SW should lead to increases in the strength of both stimulus-components. As the strength of S approaches its asymptote, the difference between the strengths of S and W should gradually diminish and finally disappear. If a child has sufficient trials before criterion is reached in Series II for the strengths of S and W to approach equality, he should choose one stimulus as often as the other in Series III. He would be classified as inconsistent. Let it be noted as an aside to be taken up presently that this analysis implies that Series II should be learned more slowly by inconsistent than by non-reversing children. Consideration of the mechanism of acquired distinctivness of stimuli[17] would lead to the expectation that mediators, *i.e.*, reversers, would learn Series II more rapidly than either of the two non-mediating groups.[18]

---

[16] It is assumed, for the present, that the removal of a positive reinforcer for an incorrect response does not alter the gross relationships between reinforcement and extinction postulated here.

[17] Miller and Dollard, *op. cit.*, 71-74.

[18] G. N. Cantor, The effects of three types of pretraining on discrimination learning in preschool children, *J. exp. Psychol.*, 49, 1955, 339-342; K. J. Norcross and C. C. Spiker, Effects of mediated associations on transfer in paired-associate learning, *ibid.*, 55, 1958, 129-134.

As a corollary to the rise in the percentage of mediators with age, there should, of course, be a decrease in the percentage of non-mediators. There seemed, at the outset of the research, no reason why the decrease should not occur for both non-mediating categories. Fig. 2 shows, however, that while there was a sharp and statistically significant drop for the inconsistent category ($p = 0.001$), there was no drop for the non-reversals, a result that bears explanation.

Despite the failure of the non-reversal category to reveal any decreasing trend, it seems reasonable to conclude that, in general, the results of Fig. 2 bear out the implication that there is a transition in the course of human development from unmediated, single-unit S–R behavior to mediated behavior, at least with reference to the acquisition of the concepts of size and brightness. They suggest further that the proportion of children who have made this transition increases in a gradual and lawful manner. It remains for further research to determine whether the same or similar relationships will obtain with other concepts.

TABLE II

NUMBER OF TRIALS TO REACH CRITERION ON SERIES I AND II AS RELATED TO CHOICE-BEHAVIOR IN SERIES III

Choice	*N*	Series I			Series II		
		M	*Mdn*	range	*M*	*Mdn*	range
Reversal	80	12.3	5.5	0–67	8.4	3.5	0–85
Nonreversal	41	31.1	15.0	0–177	14.7	6.0	1–191
Inconsistent	31	45.2	28.0	0–204	17.4	8.0	0–164

In addition to these results there are a series of observations about the training that preceded the choice-behavior and about the verbalizations that followed it that provide interesting suggestions about the nature of the mediating process and of its development. These are *ad hoc* observations and are consequently treated as such. That is, they are not offered as confirmed results supported by statistical significance, but rather as suggestive observations requiring further experimental verification.

Table II shows how the speed of learning differed in Series I and II for the children whose later behavior in Series III placed them in one of the three categories of choice. Because of the skewness of the distributions both the means and medians are presented. The trend of results is, however, the same for either measure. In both series, the children who mediated (reversing group) learned fastest. The non-reversers were next and the inconsistent slowest. This is precisely the order predicted for learning in Series II. Moreover, there is partial corroboration of these results in

Buss' study in which he found that college students when presented with a similar task fell 'unequivocally' into either the reversing or non-reversing category.[19] When he compared their initial learning, he found that there was no difference in Series I between the two groups but that Series II was acquired more rapidly by the reversers than by the non-reversers. Thus, the results of learning in Series II are consonant with the theoretical framework and with previously reported results.

The explanation of the order found for Series I, which is the same as in Series II, does not flow as readily from the present analysis nor is it corroborated by Buss' study. There is, however some indirect corroboration in the previously cited study with kindergarten children,[20] where it was found that fast learners of the initial series reversed more readily than slow learners, which is, of course, what was found for Series I. One possible explanation of these results is that reversing children, *i.e.* children who mediate in Series II, may learn Series I more rapidly because they also mediate there. They may bring to Series I relevant mediational mechanisms, such as covert verbalizations, which could enable them to sort out the two available concepts quickly and guide their overt behavior accordingly. Such theorizing implies that the increasing ability to learn associated with increasing age is related to the increasing availability of relevant mediators. College students are generally likely to have in their repertoire the mediators that produce rapid learning in the type of problem presented by Buss. Buss reports very rapid initial learning (medians of 2 trials for reversers and 3 trials for non-reversers) and so there is little or no difference between the categories in the initial series.

The last set of data to be presented is drawn from the answers given by the children to questions about what the 'winner' looked like after Series III was completed. These questions were answered in the presence of the last pair of stimuli with which they had been previously presented. The children were classified according to the descriptions they offered. The responses used in this classification were as follows: *Verbalized correct dimension* refers to responses which mentioned either one or both of the dimensions that had apparently guided *S*'s behavior in Series II and III, *e.g.* of a child who had been responding to brightness in the preceding series and used the terms 'black' or 'white.' *Verbalized incorrect dimension* refers to responses of children who described the dimension that was inconsistent with their overt choice, *e.g.* of a child who had been responding to brightness but limited his verbal description to size. *Irrelevant or no verbalization* included responses of three groups of children; of those who gave no response, of those who pointed at the stimulus but provided no spoken comment, and of those whose spoken comments did not include either dimension on which the stimuli differed. Examples of the last group are children who said, "It's the square one," even though all stimuli are square. One child said it was "the camel."

---

[19] Buss, *op. cit.,* 1956, 162-166.

[20] Kendler and Kendler, *op. cit.,* 56-60.

Table III presents the categories of verbal description in terms of *S*'s previous choice-behavior. When these results are interpreted, it should be understood that they describe behavior *after* the learning and test-series and under circumstances that maximize the production of a verbal response in terms of at least one of the dimensions. One of the most interesting observations is that despite the pressures to respond generated by *E*'s persistent questions, with the stimulus-cards in view, 42% of the inconsistent children failed to produce any relevant verbalization. If verbalization is important for the mediating process, then it would follow that non-mediators would be relatively inarticulate. By the same token, mediating children should produce a relatively large proportion of verbal comment that was relevant to their previous performance, and that is exactly what Table III shows. If the pattern is clear for the reversers and inconsistent children, the non-reversers present a more complex picture. From Table III two things may be gleaned about this group; one is that an overwhelming proportion produced descriptions of the stimuli in terms of at least one of the manipulated dimensions. This proportion was just as large as the reversing group, suggesting that, under pressure to do so, they could verbally describe the stimuli as well as the mediators. But the other observation is that these verbalizations were less frequently relevant to their previous behavior than was true for the reversing *S*s.

TABLE III

PERCENTAGE OF *S*S GIVING VARIOUS DESCRIPTIONS AS A FUNCTION OF THEIR CHOICE IN SERIES III

Choice	Descriptions (verbalizations)		
	correct	incorrect	not relevant
Reversal	84.8	7.6	7.6
Non-reversal	66.7	25.6	7.7
Inconsistent	——57.7*——		42.3

* If the behavior was categorized as inconsistent, neither dimension could be considered correct; therefore, mention of either dimension was combined and placed between the two columns to make clear its special character.

One tentative way to tie these observations together and simultaneously throw further light on the failure of the proportion of non-reversing children to decrease with age is to propose that reversing, non-reversing, and inconsistent choice-behavior represents a three-stage heirarchy of development. Reversal-choice reflects the highest level; here covert verbal responses occur during training and they also serve to mediate behavior. Non-reversal constitutes an intermediate level at which covert verbal responses can occur and sometimes do, but they either occur rather late in the learning or they do not necessarily or readily mediate overt behavior. The most primitive level is characterized by little or no covert response and is manifested in inconsistent behavior. With increasing *CA* more and more children reach the highest level (*i.e.* reversal) and fewer and fewer are left at the lowest level (*i.e.* inconsistent), but at each age tested the proportion in transition between the two extreme levels (*i.e.* non-reversals) tends to be constant (see Fig. 2).

Such an analysis would lead to the expectation that the proportion of children who verbalized correctly would increase with age; that the proportion of children who gave either no verbalizations or irrelevant ones would decrease with age and that the proportion of children verbalizing the incorrect dimension would not change much. Fig. 3 which presents the data on verbalization in terms of chronological age shows a considerable correspondence between expectation and results. It also shows a striking similarity to the choice-behavior plotted in Fig. 2, a similarity that occurs despite the fact that the children who comprise each set of parallel developmental trends are not identical. For example, the curve of correct verbalization in Fig. 3, which parallels the curve of reversal-choice in Fig. 2, includes 67% of the non-reversing children as well as 85% of the reversing children. Thus, although these results cannot point to a perfect relationship between verbalization and choice-behavior, the similarity of trends certainly suggest that the development of the mediational process is intimately related to the development of the ability to relate words to actions.

There is one more suggestive result yielded by the data on verbalization which may simultaneously help to explain why such a high proportion of non-reversing children verbalized correctly, why the curve of reversal approaches such a low asymptote and why the curve of non-reversal shows no sign of decreasing even at the ripe old age of 10 yr. During classification of the verbal behavior, it was observed that a few children, particularly among the older groups, described the 'winner' accurately by mentioning both dimensions, *e.g.* "The big white one." To investigate this observation, the *S*s grouped under *Verbalized Correct Dimensions* in Fig. 3

were subdivided into *S*s who described one dimension and *S*s who described both, with the result as presented in Fig. 4. It can be seen that the curve for one dimension, like the curve for reversers, reaches its highest point at a little over 60%. It can also be seen that the proportion of children who describe both dimensions is 0 at 3 yr. of age and increases slowly to 25% at age 10 yr., which implies, reasonably enough, that there is a developmental aspect to the number of simultaneous mediating responses a child can handle. If this is the case, then it may be that at the upper end of the scale of chronological age, in a situation as simple as Series II, non-reversing behavior does not necessarily result from a primitive single-unit S–R association. It may represent a more advanced process than reversal-choices since it could arise from the ability to integrate more than one mediating process.

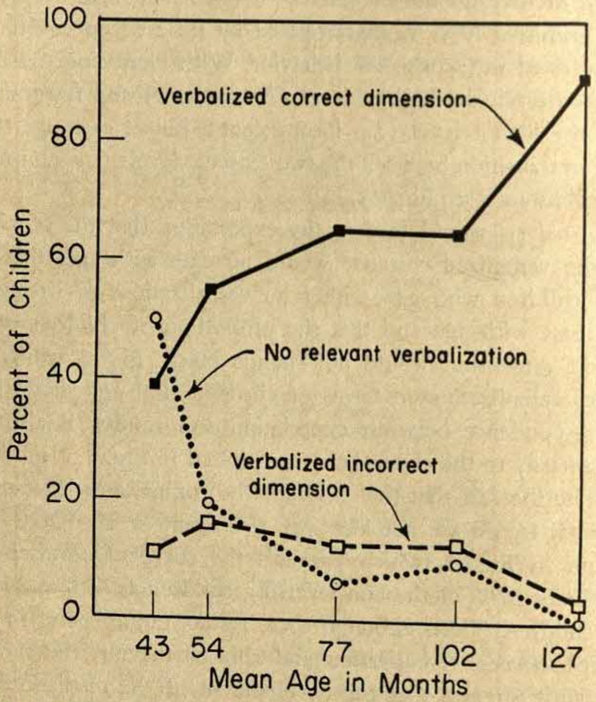


FIG. 3. PERCENTAGE OF CHILDREN AT EACH AGE-LEVEL RESPONDING IN EACH VERBAL CATEGORY

The foregoing is a somewhat more precise way of saying that the task may have been too easy for the older children and that they complicated it for themselves. If this analysis is correct, then the failure of the curve of reversal to rise above 62% and the curve of non-reversal to drop below 28% may be due to the ever-present difficulty in developmental research; that of providing one task which is easy enough for the lower end of the scale and yet difficult enough to pose the proper challenge at the upper end. What this amounts to is that the present task was capable of differentiating among the various age-levels tested, but may have attenuated the differences at the upper age-levels. If so, it may not be too large a price to pay for clues to the understanding of how this attenuation occurs, especially if it uncovers another process and suggests methods for its measurement or control in future research.

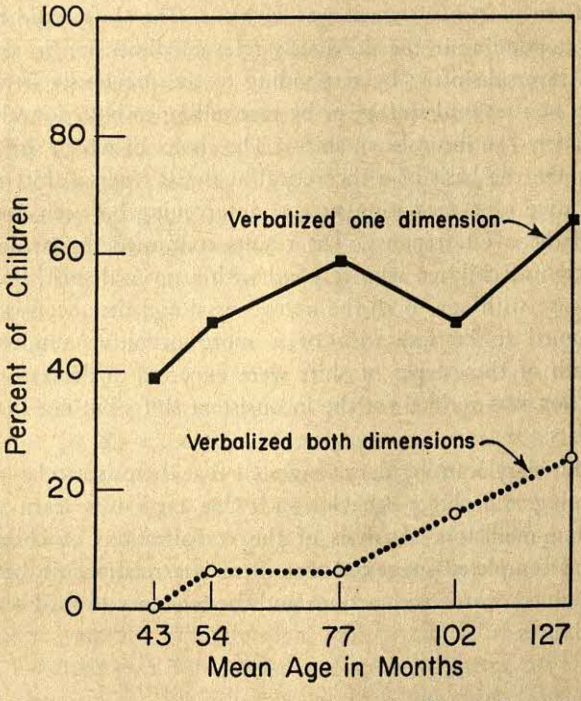


FIG. 4. PERCENTAGE OF CHILDREN AT EACH AGE-LEVEL RESPONDING IN EACH OF THE SUBCATEGORIES OF CORRECT VERBALIZATION

SUMMARY

The study was undertaken to test the implications of previous research to the effect that children below 5 or 6 yr. of age tend to behave predominantly on a single-unit S–R basis and that with increasing chronological age an increasing proportion of children behave in a mediated manner. Children at 5 age-levels; namely 3, 4, 6, 8, and 10 yr., were presented with an initial discriminative task of concept-formation that included a relevant and an irrelevant stimulus-dimension. After criterion they were presented with a shift in which the same stimuli were employed but the pattern of reinforcement was reversed. The shift could be accomplished by responding to the previously relevant dimension in an opposite manner (a reversal-shift) ; by responding to the previously irrelevant dimension (a non-reversal shift) ; or by responding to both stimulus-aspects indiscriminately (an inconsistent shift). The choice of which shift he made lay with *S*. On the basis of a theoretical analysis, reversal-shift was taken to be indicative of a mediating process intervening between the external stimuli and the overt response. The results confirmed the prediction that the proportion of children who respond with a reversal-shift, *i.e.* mediate, would increase with age. Both the non-reversal and the inconsistent shifts were presumed to be indicative of a more primitive single-unit S–R process. Both of these types of shift were expected to decrease with age. The prediction was verified for the inconsistent shifts but not for the non-reversal shifts.

The preliminary learning data suggested that there is a relationship between learning and choice-behavior such that mediators learn more rapidly than non-mediators. Analysis of the verbalizations of children after the task was completed suggested that there is a relationship between the ability to connect words with actions and the tendency toward reversal, *i.e.* mediated choices.

THE APPARENT SIZE OF AFTER-IMAGES IN DELUSIONAL AND NON-DELUSIONAL SCHIZOPHRENICS

By ALAN M. HARTMAN, VA Hospital, Palo Alto, California

Recent investigations of the perception of constancies by differentiated groups of schizophrenics suggest that these persons differ from normal in terms of perception of size- or distance-constancy, as would be expected on the basis of assumed underlying personality-structure.[1] The reliability of group-differences demonstrated in these studies is, however, never high enough to encourage the use of constancy-perception for individual diagnosis, nor has there been a systematic attempt to partial out critical sources of variance in the perception of the constancies, *e.g.* ability of the subjects (*Ss*) to use comparison-methods of measurement, maintenance of motivation to perform in the experimental situation, the role of manifest anxiety, and the possibility that ocular factors might account for the group-differences reported in such studies.

The perception of after-images may be assumed to be a rough converse of the perception of size-constancy, or as a function of retinal areas of stimulation, depending upon how the phenomena are measured and the cues available in the experimental situation.[2] If, however, the two approaches are combined, *i.e.* more than one method of measuring or of comparing the size-changes of the after-image (*AI*) is employed, then a situation results in which the role of ocular, motivational, and measurement-ability can be assessed. Physical measurement of the changes in the size of the *AI* with distance—Emmert's Law—can be used to assess the role of ocular variables, *S*'s attention, and his ability to use the various systems of measurements. Measurement of apparent size of the after-images provides a check on the quality and quantity of the cues of distance necessary for *S* to report changes in size as well as a check on his ability to use the methods of measurement. It is assumed here that the physical and apparent size of *AIs* may be used as a dependent variable in much the same manner as size-constancy, but with the added advantage that more sources

---

* Received for publication November 7, 1960.

[1] H. L. Raush, Perceptual constancy in schizophrenia, *J. Person.*, 21, 1952, 176-187; T. E. Weckowicz, Size constancy in schizophrenic patients, *J. ment. Sci.*, 103, 1957, 475-486; T. E. Weckowicz, R. Sommer, and R. Hall, Distance constancy in schizophrenic patients, *ibid.*, 104, 1958, 1174-1182; and G. W. Granger, Personality and visual perception: A review, *ibid.*, 69, 1953, 8-43.

[2] Ward Edwards, Apparent size of after-images under conditions of reduction, this JOURNAL, 66, 1953, 449-455; F. A. Young, The projection of after-images and Emmert's Law, *J. gen. Psychol.*, 39, 1948, 161-166.

of variance can be assessed than is possible in the orthodox approach to size-constancy.

The primary purpose of the present study was to evaluate the following hypotheses:

Hypothesis 1. The chronic, delusional, schizophrenic patient demonstrates a higher degree of size-constancy than non-delusional schizophrenics or normal control *Ss*, and therefore will be less likely to respond as would be predicted on the basis of Emmert's Law.

Hypothesis 2. There is no difference between the delusional schizophrenic and normal *Ss* in terms of the physical bases of Emmert's Law. That is, the size of the *AI* on the *S*'s retina is not affected in a delusional schizophrenic and hence there is no ocular basis for the schizophrenic's deficit in the dynamics of *AI*-projection.

Hypothesis 3. There is no difference between the delusional schizophrenic, non-delusional schizophrenic, and the normal *S* in the ability to make use of the physical outlining or comparative methods of measurement.

Hypothesis 4. There will be a correlation between manifest anxiety—Taylor Scale—and the degree of size-constancy demonstrated by the delusional, non-delusional, and normal *Ss*.

Since this study utilized measures of projected after-images in a reduced cue-situation a secondary purpose of the study was to replicate Edwards' findings that there is a converse relationship between size-constancy and the apparent sizes of after-images.[3]

Procedure: (1) Subjects. Four groups of *Ss* were used in this study: a control group; two experimental schizophrenic groups; and a non-schizophrenic group.

(a) Control. The control group of *Ss* (Group C) was composed of 10 men and 5 women, all staff members or trainees in the psychological unit at the Hospital. They were aware of the nature of Emmert's Law and were highly trained in the use of the measuring devices.

(b) Experimental. Two groups of schizophrenic *Ss* (Groups E_1 and E_2) were used. Group E_1, the delusional schizophrenic group, contained 15 patients who were characterized by their ward psychologists as having a long history of recurrent, systematic delusions. The breadth and intensity of their delusions were demonstrated in many areas of their interactions within their environment and their interpersonal relationships. Their mean age was 37.5 yr., *SD* 5.8 yr., and their mean years of hospitalization was 4.6, *SD* 2.8. Group E_2, the non-delusional schizophrenic group, contained 15 men who were considered by their ward psychologists as comparable to the delusional patients in all respects other than the severity of their delusional thinking. The mean age of this group was 36.07 yr. with a *SD* of 6.32 yr. Their mean years of hospitalization was 3.93, with a *SD* of 2.51. Additional criteria for the selection of the *Ss* for the schizophrenic groups were as follows: not lobotomized, not so aggressive, resistive, or confused as to be untestable; hospitalized for a nervous disorder for more than one year; and between the ages of 25 and 51 yr.

(c) Non-schizophrenic. A fourth group (Group N) composed of 15 men with no history of neuropsychiatric hospitalization, were asked to volunteer for a study of after-images. Their mean age was 36.87 yr. with a *SD* of 6.38 yr.

(2) Apparatus. The apparatus duplicated Edwards' with the exception that two

---

[3] Edwards, *op. cit.*, 449.

additional systems of measurement were added.[4] The left and right fields were set across two tables in a well-ventilated, darkened room. The units, described below, were painted flat black inside and out, and black flock paper was added to reduce reflection (Fig. 1).

(a) The left-hand tunnel consisted of a trapezohedral shaped box and contained projection Screens B and C (43 and 90 in. from *S* respectively) and the outlining rods. The base of the tunnel was 14.5 x 22 in., and the apex was 4.5 x 6.5 in.

(b) The right-hand tunnel (7.5 x 5.5 x 114 in.) contained the comparison-square which moved back and forth on a toy electrical track.

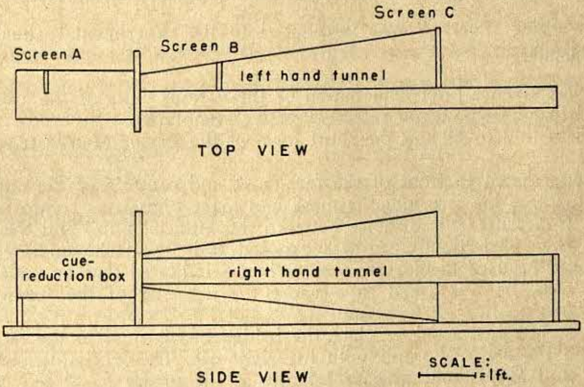


FIG. 1. SCHEMATIC DIAGRAM OF THE EXPERIMENTAL TUNNELS, PROJECTION-SCREENS, AND THE CUE-REDUCTION BOX

(c) The cue-reduction box (11 x 12 x 12 in.) connected the left- and right-hand fields and housed the face-mask, the third projection-screen (6.75 in from *S*), and the cue-reduction mask. The cue-reduction mask had a rectangular opening of 3 x 7 in.

(d) The face-mask was fastened to the end of the cue-reduction box and consisted of a 2-in. band of aluminum which was so cut and taped that *S* could comfortably look through a 3-mm. artificial pupil set in line with the common partition of the right- and left-hand fields. This unit insured that *S* would maintain a constant position and distance in line with the tunnels.

(e) The *AI* induction-square consisted of a 1-in. sq. piece of clear plastic with a black fixation-point at its center. The square was set into a mask which, when lowered into position 26 in. from *S*, blocked his view through both tunnels. The *AI*-induction square was illuminated from behind by a 150-w. flood-lamp set 43 in. from *S*.

(f) The outlining system of measuring consisted of two rods ⅛-in. in diameter mounted directly behind Screen B (42 in. from *S*) and two rods, ¼-in. in diameter mounted directly behind Screen C (90 in. from *S*). Each set of rods was so attached to control wires that *S* was able to operate them from his position behind the face-mask.

(g) The moving comparison-square was a 1-in. sq. of ground glass set in the front of a light-tight box. The box contained a ½-w. neon bulb and the whole assembly was attached to a toy electrical flat car. *S* could vary the position of the box between 14 and 104 in. from the face-mask through a string and pulley system.

---

[4] Edwards, *op. cit.*, 450-451.

of variance can be assessed than is possible in the orthodox approach to size-constancy.

The primary purpose of the present study was to evaluate the following hypotheses:

Hypothesis 1. The chronic, delusional, schizophrenic patient demonstrates a higher degree of size-constancy than non-delusional schizophrenics or normal control *Ss*, and therefore will be less likely to respond as would be predicted on the basis of Emmert's Law.

Hypothesis 2. There is no difference between the delusional schizophrenic and normal *Ss* in terms of the physical bases of Emmert's Law. That is, the size of the *AI* on the *S*'s retina is not affected in a delusional schizophrenic and hence there is no ocular basis for the schizophrenic's deficit in the dynamics of *AI*-projection.

Hypothesis 3. There is no difference between the delusional schizophrenic, non-delusional schizophrenic, and the normal *S* in the ability to make use of the physical outlining or comparative methods of measurement.

Hypothesis 4. There will be a correlation between manifest anxiety—Taylor Scale—and the degree of size-constancy demonstrated by the delusional, non-delusional, and normal *Ss*.

Since this study utilized measures of projected after-images in a reduced cue-situation a secondary purpose of the study was to replicate Edwards' findings that there is a converse relationship between size-constancy and the apparent sizes of after-images.[3]

Procedure: (1) Subjects. Four groups of *Ss* were used in this study: a control group; two experimental schizophrenic groups; and a non-schizophrenic group.

(a) Control. The control group of *Ss* (Group C) was composed of 10 men and 5 women, all staff members or trainees in the psychological unit at the Hospital. They were aware of the nature of Emmert's Law and were highly trained in the use of the measuring devices.

(b) Experimental. Two groups of schizophrenic *Ss* (Groups E_1 and E_2) were used. Group E_1, the delusional schizophrenic group, contained 15 patients who were characterized by their ward psychologists as having a long history of recurrent, systematic delusions. The breadth and intensity of their delusions were demonstrated in many areas of their interactions within their environment and their interpersonal relationships. Their mean age was 37.5 yr., *SD* 5.8 yr., and their mean years of hospitalization was 4.6, *SD* 2.8. Group E_2, the non-delusional schizophrenic group, contained 15 men who were considered by their ward psychologists as comparable to the delusional patients in all respects other than the severity of their delusional thinking. The mean age of this group was 36.07 yr. with a *SD* of 6.32 yr. Their mean years of hospitalization was 3.93, with a *SD* of 2.51. Additional criteria for the selection of the *Ss* for the schizophrenic groups were as follows: not lobotomized, not so aggressive, resistive, or confused as to be untestable; hospitalized for a nervous disorder for more than one year; and between the ages of 25 and 51 yr.

(c) Non-schizophrenic. A fourth group (Group N) composed of 15 men with no history of neuropsychiatric hospitalization, were asked to volunteer for a study of after-images. Their mean age was 36.87 yr. with a *SD* of 6.38 yr.

(2) Apparatus. The apparatus duplicated Edwards' with the exception that two

---

[3] Edwards, *op. cit.*, 449.

additional systems of measurement were added.[4] The left and right fields were set across two tables in a well-ventilated, darkened room. The units, described below, were painted flat black inside and out, and black flock paper was added to reduce reflection (Fig. 1).

(a) The left-hand tunnel consisted of a trapezohedral shaped box and contained projection Screens B and C (43 and 90 in. from *S* respectively) and the outlining rods. The base of the tunnel was 14.5 x 22 in., and the apex was 4.5 x 6.5 in.

(b) The right-hand tunnel (7.5 x 5.5 x 114 in.) contained the comparison-square which moved back and forth on a toy electrical track.

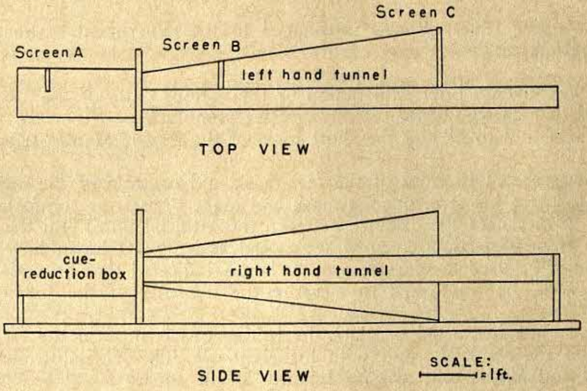


FIG. 1. SCHEMATIC DIAGRAM OF THE EXPERIMENTAL TUNNELS, PROJECTION-SCREENS, AND THE CUE-REDUCTION BOX

(c) The cue-reduction box (11 x 12 x 12 in.) connected the left- and right-hand fields and housed the face-mask, the third projection-screen (6.75 in from *S*), and the cue-reduction mask. The cue-reduction mask had a rectangular opening of 3 x 7 in.

(d) The face-mask was fastened to the end of the cue-reduction box and consisted of a 2-in. band of aluminum which was so cut and taped that *S* could comfortably look through a 3-mm. artificial pupil set in line with the common partition of the right- and left-hand fields. This unit insured that *S* would maintain a constant position and distance in line with the tunnels.

(e) The *AI* induction-square consisted of a 1-in. sq. piece of clear plastic with a black fixation-point at its center. The square was set into a mask which, when lowered into position 26 in. from *S*, blocked his view through both tunnels. The *AI*-induction square was illuminated from behind by a 150-w. flood-lamp set 43 in. from *S*.

(f) The outlining system of measuring consisted of two rods ⅛-in. in diameter mounted directly behind Screen B (42 in. from *S*) and two rods, ¼-in. in diameter mounted directly behind Screen C (90 in. from *S*). Each set of rods was so attached to control wires that *S* was able to operate them from his position behind the face-mask.

(g) The moving comparison-square was a 1-in. sq. of ground glass set in the front of a light-tight box. The box contained a ½-w. neon bulb and the whole assembly was attached to a toy electrical flat car. *S* could vary the position of the box between 14 and 104 in. from the face-mask through a string and pulley system.

---

[4] Edwards, *op. cit.*, 450-451.

The illumination of the comparison-square was 0.51 apparent foot-candles at 104 in., and 0.99 at 14. in.

(b) The comparison-square consisted of a 3-in. piece of ground glass which was covered by a square diaphragm. The diaphragm and glass assembly was fixed to a 6-in. light-tight cube which contained a ¼-w. neon bulb. Illumination of the square was 1.95 apparent foot-candles at the full open position—the width of the square could be varied continuously from 0 to 45 mm. An aluminum frame was so mounted above the diaphragm that the 6-mm. artificial pupil in the apex was 5.25 in. from the ground glass.

The illumination of the projection-screens was as follows: Screen A (6.75 in. from *S*), 0.14 apparent foot-candles; Screen B (43 in. from *S*), 0.34 apparent foot-candles; and Screen C (90.5 in. from *S*), 0.22 apparent foot-candles.

(3) Pre-training. Every *S* was introduced to the experimental situation in the following manner:

(a) He was given a brief explanation of the various parts of the apparatus, and was assured that he would not receive aversive stimulation.

(b) He was told to fill out the short form of the Taylor Manifest Anxiety Scale (*MAS*).

(c) He was seated in front of the face-mask and controls of the apparatus and was asked to place his face into the mask and make himself as comfortable as possible while *E* measured the distance between the artificial pupil and the *S*'s cornea.

(d) The near projection-screen was exposed and *S* was shown how the outlining rods could be used to measure size. When *S* indicated that he understood how to use the rods, he was asked to measure the diameter of the fixation-point on Screen B.

(e) *S* was then instructed to place the bridge of his nose in the face-mask and peer through the artificial pupil with his right eye. The *AI* induction-square was illuminated and *S* was told to gaze intently at the square for 9 sec. When *S* reported the *AI*, he was asked to gaze at a piece of paper held in front of him by *E* who slowly moved the paper toward *S* until *S* reported that the apparent size of the *AI* decreased.

(4) Training. After the pre-training period, *S* was drilled in the following tasks.

(a) Outlining rods. The *AI* was induced in the manner described above and *S* was told to use the appropriate control-knobs to measure the width of the *AI* that he saw on the screen.

(b) Moving comparison-square. *S* was shown the orange square in the right-hand tunnel and was asked to report what he saw when the square was moved forward or backward. He was told that the comparison-square was to be used to indicate the size of the *AI* and that he should keep the *AI* on the grayish-green, left-hand projection-screen and not to permit it to drift over and cover the orange comparison-square. The comparison-square was set at 104 in. at the start of every trial.

(c) Diaphragm comparison-square. *S* was shown how the diaphragm-square could be used to indicate size. He was told just to look at the *AI* as it appeared on the projection-screens and try subjectively to estimate its size. When he thought that he could estimate the size of the projected *AI* he was told to close his right eye, open his left eye, and make the diaphragm-square the same size as the projected *AI*.

Every *S* was given four outlining rod-trials at each of the planes of projection set at 43 and 90 in. Every *S* was given four trials with the moving-square and four trials with the diaphragm-square for each of the planes of projection set at 6.75 and 90 in. Furthermore, every *S* received two additional trials with both the diaphragm and the moving-square to match the *AI* at the 43-in. projection-distance. The order of the trials at different distances of the projection-screen was randomized.

Results. The results of the trained *S*s (Group C) in measuring with the outlining rod did not differ significantly from the values predicted on the basis of Emmert's Law, but none of them reported that he had observed a change in the size of the *AI.* These *S*s did not, however, use the comparison-square methods to indicate changes in the size of the *AI* when the distance of the projection-screens were set at 43 in. and then at 90 in.

The results of adding monocular cues to distance to determine the cues necessary for the *S*s to perceive variations in the size of the *AI* were as follows:

(1) When identical figures, the numeral '4' for example, were pasted on both screens, the *S*s reported that one projection-screen was farther away from the face-mask than the other; they were, however, still unable to report a size-change in the *AI.*

(2) Screen B (43 in.) was placed halfway out of the tunnel so that the *S*s could rapidly alternate their fixation from one screen to the other. When this was done, they still reported that the apparent size of the *AI* remained the same.

(3) When the *S*s projected their *AI*s on the screens, set completely outside of the left-hand tunnel, but at the same distances from the face-mask, none of them was able to report a change in the size of the *AI* at the distances at which the projection-plane was varied (43 to 90 in.).

(4) When the constancy of the ratio of the area of the *AI* to the area of the projection-screen was destroyed by moving the large projection-screen back and forth outside the tunnel to the 43- and the 90-in. distances, the *S*s were able to report marked changes in the size of their projected *AI.*

(5) When the projection-screens were set at distances of 6.75 and 90 in., the *S*s were then able to report reliably a change in the size of the *AI* as would be expected if Emmert's Law had a phenomenological correlate. This suggests that convergence influences the apparent size of the *AI* when the projection-distance is closer to *S* than the stimulus inducing the *AI.*

The non-delusional schizophrenics (Group E_2) used the moving-square method in a manner which suggests that they were able to distinguish slight changes in the size of the *AI* when the projection-screens were set at 43 and then at 90 in. A *t*-test of the differences in the perception of the size of the *AI* at these distances was significant at better than the 5% level of confidence. All between-group comparisons of ability to perceive changes in the size of the *AI* when the projection-screens were set at 43 and then at 90 in. were, however, non-significant. These findings, and the fact that all groups used each method of measurement reliably, confirm Hypothesis 3 that there would be no difference in the ability of the three groups to measure the size of the *AI*, independent of the systems of measurement. When cues of the size of *AI* were absent, the groups did not differ in their perception of the apparent size of the *AI*.

The results confirm Hypothesis 1 that delusional and non-delusional personality affects or contaminates the perceptual processes. The assumption of homogeniety of variance was supported. The data for Groups E_1, E_2, and N were, therefore, submitted to a mixed-type of analysis of variance. The results showed a significant effect due to the methods of measurement and a significant interaction between the groups and the effect of the methods. This interaction was interpreted to indicate that the groups perceived differently when at least one method of measuring the size of the *AI* was used. A series of *t*-tests showed that Groups E_1, E_2, and N differed from each other in terms of size-estimates of *AI*s obtained with the moving comparison-square.

Hypothesis 2 that there is no ocular basis for the schizophrenic's deficit in the *AI* project was confirmed. A mixed-type analysis of variance of the difference between obtained—outlining rod measures—and expected values of the size of the *AI* indicates that there were no between-group-differences in the discrepancies of observed and theoretical values of the *AI*, nor was there an interaction between the groups and the distances of the projection.

Hypothesis 4 that manifest anxiety scores will be correlated with constancy estimates was not substantiated. Manifest anxiety, as measured by the Taylor Scale, was correlated with the degree of constancy for the mean scores of the different groups. The correlations based on individual scores were not, however, significant. The non-schizophrenic groups (Groups C and N) showed the least anxiety; the delusional group (Group E_1) was intermediate; and the non-delusional schizophrenics (Group E_2) were highest. The assumption of homogeneity of variance was upheld and the analysis indicated that one or more of the groups varied significantly from the others at the 1/10% level of confidence. A series of *t*-tests showed that all groups differed significantly from the others at better than the 2.5% level of confidence.

Discussion. The Edwards-Boring phenomenological interpretation of Emmert's Law assumes that, if all cues to distance are removed from a viewing situation, the size of the *AI* would depend mainly on its retinal size which remains constant under reduced cue-conditions.[5] *S* should not, therefore, perceive a change in the size of the *AI* as the distance of projection varies. Furthermore, the phenomenological size of the *AI* is considered to be the converse of size-constancy because retinal size remains constant and *S* perceives a change in the size of the *AI* as the distance of the projection-plane varies. One implication of this interpretation of Emmert's Law is that cues of distance affect the perceived size of the *AI*. The introduction of cues of distance or depth should, therefore, change the perceived size of the *AI* and Emmert's Law should be considered as the converse of size-constancy. The results of this study support Edwards' findings,[6] but more important, the present study provides information about the cues necessary for the perception of changes in the apparent size of the *AI*s. In this study, monocular cues of distance were available, but no *S* was able to report that his *AI* had changed as the distance of the projection-screens varied. If the perception of changes in the size of the *AI* depends on cues to the distance of the projection-screens, then the *S*s should have been able to report apparent changes in size.

---

[5] E. G. Boring, Size constancy and Emmert's Law, this JOURNAL, 53, 1940, 293-295.

The present study indicates that Emmert's Law can be verified without *S*'s awareness of the changes in the size of the *AI*, but *S* is able to perceive changes in the apparent size of the *AI* only when the retinal areas of the *AI* and the projection-background vary to some extent.

Raush found that the tendency for paranoid schizophrenics to perceive size-constancy was most marked under dim illumination.[7] That is, for his three groups, paranoid schizophrenic, undifferentiated schizophrenic, and normals, the trend toward group-differences in scores of over-constancy was most significant when paranoid schizophrenics and normals were compared. Similarily, in the present study, a significant difference was found between paranoid schizophrenics (Group E_1) and the normals (Groups C and N) in the perception of apparent sizes of *AI*s, but in the present study, two sources of confounding may be rejected. First, the possibility of a physical basis for a deficit in perceptual function has been ruled out because the schizophrenics were capable of measuring the size of the projected *AI* in accordance with Emmert's Law even though in most cases they were not aware of apparent changes in size. Secondly, the data reported here indicate that all of the *Ss* were able to use the methods of measurement equally well, and hence the elimination of these two factors strengthens the conclusion that general personality variables are related to variability in basic perceptual processes.

---

[6] Edwards, *op. cit.*, 165f.

[7] Raush, *op. cit.*, 184-186.

The three experimental groups were sharply differentiated by the Taylor *MAS*, and the order of the group means correlated with the order of the constancy scores. Correlations based on the *N*, however, were not significant.

The *MAS* data served as an excellent valadity-check on the selection procedures. The groups were chosen on the basis of extent and breadth of delusional thinking; and apparently, the selection criteria of delusional thinking did correlate with anxiety as measured by the *MAS*. The non-hospitalized group of volunteters (Group N) ranked lowest in anxiety, the unsystematized schizophrenic group (Group E_2) highest, and the systematized delusional schizophrenics (Group E_1) were intermediate in terms of *MAS* scores and constancy-scores. One tentative explanation is that certain aspects of perceptual organization are highly important variables in the control of anxiety; and hence, perceptual rigidity may be just one way, albeit inadequate, of reducing manifest anxiety.[8]

Summary. This study utilized 15 trained *S*s (10 men and 5 women); 15 chronic, delusional, schizophrenic men; 15 chronic, schizophrenic, non-delusion men; and 15 normal men. They measured their projected *AI*s at three distances in the projection-plane: (a) 6.75 in.; (b) 43 in.; and (c) 90 in. All the *S*s used three methods of measuring the size of the *AI:* (1) remotely controlled outlining rods situated at 43 and 90 in. from the projection-screens; (2) moving comparison-square; (3) diaphragm comparison-square.

The results of the study were as follows: *(1)* The *S*s did not report changes in the apparent size of the *AI* when cues of distance were added to the reduced cue-situation described by Edwards.

(2) The *S*s did report changes in the apparent size of their *AI*s when the projection-screens were set at 6.75 and 90 in. Since the ratio of the area of the *AI* to the area of the projection-screen (as measured) at the *S*'s retina) was constant for the three distances of the projection-screen, we concluded that convergence cannot be ruled out as a critical variable in the perception of changes in the size of the *AI* when the projection-plane is closer to *S* than the stimulus that induced the *AL*.

(3) The chronic delusional schizophrenic *S*s differed from the chronic non-delusional schizophrenic and the normal *S*s in their preception of the size changes of the AI, *i.e.* they were less likely to respond as would be predicted on the basis of Emmert law.

---

[8] P. McReynolds, A restricted conceptualization of human anxiety and motivation, *Psychol. Rep. Monogr.*, 2, 1956, 301-304.

(4) The delusional schizophrenic *S*s do not have a physical basis for the tendency to perceive constancy of projected *AI*s when the distance of the projection varies.

(5) There was no difference in the ability of the groups to measure the size of the *AI* with the three methods of measurement.

(6) Taylor manifest scores did not correlate significantly with *S*'s ability to perceive *AI* size-change. There were, however, significant group-differences in the *MAS* scores, and these differences were in the same order as the group differences in *AI* size-estimates.

THE QUANTIFICATION OF TONAL VOLUME

By H. S. TERRACE and S. S. STEVENS, Harvard University

Tonal volume is a subjective attribute of sound, distinct from loudness, pitch, and density. It refers to the apparent size or space-filling property: the low tones of the organ have large volume, the high tones of the piccolo have small volume. Although tonal volume had been talked about for some 50 years, it got itself established as a separate attribute only in 1934 when Stevens showed that tones of different frequency could be made to sound equal in volume by an appropriate adjustment of their relative intensities.[1] This method of mapping equal-volume contours was utilized and extended by Thomas in two later studies.[2] In his first study Thomas established the form of the equal-volume contours over wide ranges of frequency and intensity.

The present study undertakes the next step, that of measuring the quantitative dependence of tonal volume on the two stimulus-parameters—frequency and intensity. Thomas knew that the volume represented by his contours increased from right to left, as he presented them, but, as he said, "We do not know the volumic difference between the contours."[3] This position is like that of a surveyor who runs a contour line around a hill with no way of knowing the elevation above sea-level. Our purpose, therefore, was to measure the 'altitudes' of the equal-volume contours.

For this measurement we used the method of magnitude estimation,[4] a procedure that has been successfully employed with some two dozen continua to determine how sensation varies with stimulus-intensity.[5] These studies have verified the rule that sensory magnitude grows as a power function of stimulus-magnitude.

Since volume varies with frequency quite as much as it varies with

---

* Received for publication July 19, 1961. This research received support from Office of Naval Research Contract Nonr-1866(15) and National Science Foundation Grant G-10716 (Report PNR-259).

[1] S. S. Stevens, The volume and intensity of tones, this JOURNAL, 46, 1934, 397-408.

[2] G. J. Thomas, Equal-volume judgments of tones, this JOURNAL, 62, 1949, 182-201; Volume and loudness of noise, this JOURNAL, 65, 1952, 588-593.

[3] Thomas, this JOURNAL, 62, 1949, 199.

[4] Stevens, The direct estimation of sensory magnitudes: loudness, this JOURNAL, 69, 1956, 1-25; Problems and methods of psychophysics, *Psychol. Bull.*, 54, 1958, 177-196.

[5] Stevens, On the psychophysical law, *Psychol. Rev.*, 64, 1957, 153-181; The psychophysics of sensory function, *Amer. Scientist*, 48; 1960, 226-253; also in W. A. Rosenblith (ed.) *Sensory Communication*, 1961, 1-33.

intensity, it was an interesting question whether the psychophysical power-law would hold for volume. Apparently it does, but with the added feature that the exponent of the power-function varies radically with frequency.

Preliminary to the main series of experiments was a laboratory exercise carried out by R. W. Pew in 1958. Pew found that *O*s having no previous experience with volume could, without any prior training, make consistent numerical estimates of the apparent volume of a series of tones that varied in both frequency and intensity. The procedure developed by Pew was essentially the same as that followed here, and the later experiments confirm Pew's major findings. Unfortunately, a detailed comparison is not possible, because Pew used a different earphone-cushion—a circumaural type.[6]

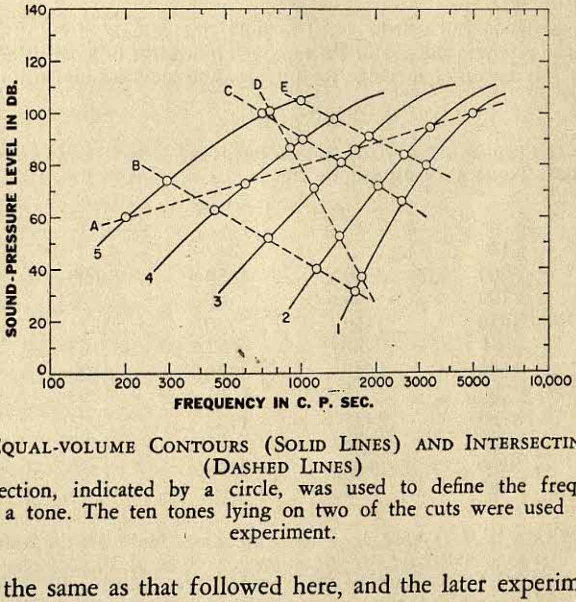


FIG. 1. EQUAL-VOLUME CONTOURS (SOLID LINES) AND INTERSECTING CUTS (DASHED LINES)

Each intersection, indicated by a circle, was used to define the frequency and intensity of a tone. The ten tones lying on two of the cuts were used in a given experiment.

PROCEDURE

The smoothed equal-volume contours presented in Fig. 8 of Thomas's paper[7] provided a basis for selecting the tonal stimuli. These contours are shown in Fig. 1, together with various dashed lines drawn across the contours. The intersections, indicated by circles, determined the frequency and intensity of the 25 tones used in the series of six experiments. Each tone is designated by a number (for the contour) and a letter (for the cut). Table I shows the values of frequency and intensity (sound-pressure level) corresponding to each intersection. The tones were generated by a beat-frequency oscillator.

---

[6] Stevens, Earphone socket, U. S. Patent 2,408,490, Oct. 1, 1946.

[7] Thomas, this JOURNAL, 62, 1949, 196.

The earphones (Permoflux PDR-8, mounted in supra-aural cushions MX-41/AR) were calibrated on a 6-cc. coupler and found to give a flat response over the range of frequencies used. No attempt was made to correct the calibration-response for the factors that enter into the real-ear response of these earphones and cushions.

In each experiment the volume of 10 tones, defined by 2 cuts in Fig 1, was judged by the method of magnitude estimation. The 10 tones were each judged twice in an irregular order that was different for each *O*. One of the two tones from the middle contour, No. 3, served as the standard in each experiment and was assigned the modulus 10. Every *O* was given written instructions as follows:

> In this experiment you will be asked to judge the volume of each of a series of tones. Volume refers to the size of the tone, *i.e.* how large or small it appears to be. Although it may be easier to judge the pitch or the loudness of the tones that you will hear, please be sure your judgments are based only on the volume of the tones. The first tone will have, by definition, a volume of 10. After you hear each succeeding tone, report the number that you feel best describes the volume of that tone—on a scale on which the volume of the first tone you will hear is 10. Are there any questions?

TABLE I

FREQUENCY (UPPER ENTRY) AND SOUND PRESSURE LEVEL (LOWER ENTRY) OF TONES DETERMINED BY THE INTERSECTIONS IN FIG. 1

Cut	Contour				
	1	2	3	4	5
A	5000	3350	1650	600	200
	100	95	86	73	60
B	1650	1150	740	455	290
	31	40	52	63	74
C	2550	2025	1450	1025	700
	66	72	81	90	100
D	1750	1425	1125	905	750
	37	53	71	87	101
E	3200	2600	1875	1350	1000
	80	84	91	98	104

Ten *O*s served in each experiment. In all, 46 *O*s were used. Slightly more than half the *O*s were graduate students in psychology who were familiar with the method of magnitude-estimation in other contexts. The others were students unfamiliar with the procedure. None of the *O*s had ever judged tonal volume.

RESULTS

The results of the magnitude-estimations are shown in Table II. The notable feature of the data in Table II is the similarity of the estimations of volume for pairs of tones taken from the same equal-volume contour. This agreement is what we should expect, of course, if Thomas's contours are essentially correct and if *O*s can indeed make valid numerical estimates of tonal volume.

Both the medians and the geometric means of all judgments were computed. The entries in Table II are geometric means, except for three instances. These are cases in which, because of one or two widely divergent judgments, the geometric means differed rather widely from the medians. In these three cases (Tones A1, A2, and E2) the medians were entered in Table II. (The corresponding geometric means were 4.4, 8.9, and 6.8.)

To determine a representative measure of variability, the median quartile deviations (in log units) for all the judgments of tones on a given contour were computed. These quartile deviations are 0.26, 0.13, 0.08, 0.12, and 0.19 for contours 1 to 5. The tendency for the variability to be least in the vicinity of the standard is not unusual in experiments on magnitude estimation. There was evidence in these experiments that the *O*s had less difficulty with the judgment of tones near the middle of the range than with tones at the extremes. Thomas noted that his *O*s had less difficulty with volume-matches when the tones were of medium frequency and intensity.[8] It should be pointed out that a large component of the variability in our experiments is presumably due to variations in the acoustic levels actually produced at the eardrums of different *O*s.

TABLE II

GEOMETRIC MEANS OF VOLUME ESTIMATIONS

Each entry is based on 20 judgments (10 *O*s). The cuts and contours refer to the designations in Fig. 1. Each experiment used 2 cuts. In each experiment the standard tone (called 10 in the instructions) was at the intersection of contour 3 and the cut listed first.

Experiment	Cut	Contour				
		1	2	3	4	5
I	C	1.6	5.1	10.1	21.5	37.5
	B	1.3	4.7	9.5	23.4	45.7
II	D	0.9	2.7	9.3	30.6	44.7
	B	1.4	2.4	7.6	27.7	37.2
III	B	1.3	3.5	10.2	17.3	29.3
	D	1.4	4.9	9.6	22.3	34.2
IV	D	1.2	4.7	9.7	26.6	40.8
	C	1.3	4.4	9.6	27.7	51.7
V	B	1.8	6.3	10.3	20.4	30.2
	E	2.5	6.0	11.7	21.2	35.8
VI	A	3.0	7.0	10.2	20.9	32.3
	D	2.6	4.2	12.1	17.2	26.7

Power-functions. To determine how volume grows with intensity, the geometric means of all the values on each contour were computed. For

---

[8] Thomas, this JOURNAL, 62, 1949, 182-201.

Contours 1 to 5 these values are 1.8, 4.8, 10.4, 24.3, and 39.2. Knowing the approximate subjective value of each contour in Fig. 1, one can read off, for a given frequency, the sound-pressure level corresponding to a given value of volume. These vertical cuts through the contours yield the points plotted in the logarithmic coördinates of Fig. 2.

Two features of the volumic attribute become immediately apparent from the data in Fig. 2. Since the values for a given frequency approximate a straight line in the log-log plot, it follows that volume grows as a power-function of sound-pressure, with a different slope (exponent) at each frequency. The other feature is a convergence of all the psychophysical functions toward a common point at about 140 db. This convergence confirms the early deduction that Stevens made from his equal-volume contours to the effect that "at every high sensation-levels the volume of all tones would be equal."[9] Since the convergence was so obvious in the basic data, the lines fitted to the points in Fig. 2 were made to pass

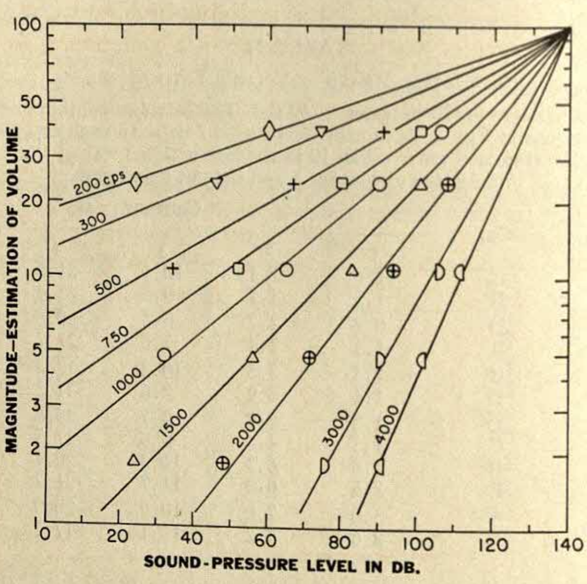


FIG. 2. PSYCHOPHYSICAL FUNCTIONS RELATING VOLUME TO INTENSITY
The points represent the values determined by vertical cuts through the contours of Fig. 1. The ordinate values were determined by a geometric averaging of all magnitude estimations relating to a given equal-volume contour.

---

[9] Stevens, this JOURNAL, 46, 1934, 404.

through the common point. As can be seen, this procedure does very little violence to the data.

The exponent as a function of frequency. The fact that the volume-functions converge to a common point provides an opportunity for another test of the general consistency of the data obtained in these experiments. For each of the 25 tones determined by the intersections in Fig. 1, the geometric mean of all the estimates in all relevant experiments was computed. This value of volume, taken in conjunction with the value at the point of convergence, determines a slope in a plot similar to Fig. 2.

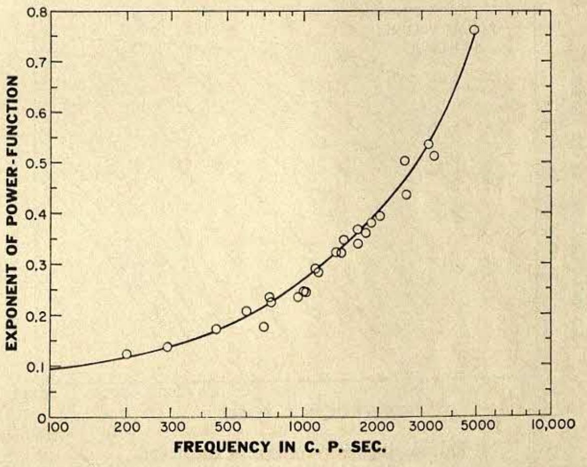


FIG. 3. THE EXPONENT OF THE POWER FUNCTION FOR VOLUME AS A FUNCTION OF FREQUENCY

A value of the exponent was determined for each of the 25 tones used in the six experiments. This procedure provides a severe test of the consistency of the data.

The slopes thereby determined give the exponents as plotted in Fig. 3. In this manner, we have let the estimations of magnitude of each tone determine the value of an exponent, and the interesting question is whether the values thus determined follow a consistent function of frequency. The answer in Fig. 3 is plain enough: it is not difficult to draw a simple monotonic function through the 25 points.

It is interesting that, according to Fig. 3, a tone of about 3750 c.p.sec. grows in volume with an exponent of 0.6. This is the same exponent that governs the growth of loudness.

The next step is to reconstruct the equal-volume contours, thereby completing the circle. For this purpose we read the exponent from Fig. 3, and we assume that the volume-functions all converge at 140 db. At this point it becomes advisable to define a unit of volume: 1 *vol* is the apparent volume of a tone of 1000 c.p.sec. at a sound-pressure level of 40 db. re 2×10^{-5} newton per square meter. From this definition, plus the slope (exponent) of the volume-function at 1000 c.p.sec., we determine that the point of convergence at 140 db. has a volume of 22 vols.

The calculated equal-volume contours are shown in Fig. 4. Although they represent a considerable extrapolation beyond any values that have been experimentally determined, their general form is consistent with all that is known about volume. In particular, the contours in Fig. 4 are quite consistent with the equal-volume measurements by Thomas.

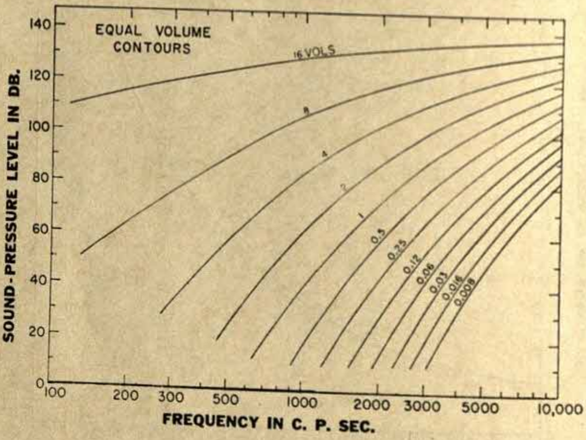


FIG. 4. EQUAL-VOLUME CONTOURS
These contours were constructed with the aid of the exponent in Fig. 3, together with the convergence-point in Fig. 2.

Volume vs. frequency. As already noted, volume is a joint function of frequency and intensity. It is of interest, therefore, to show how volume varies with frequency for a given sound-pressure level. Horizontal cuts through the contours of Fig. 4 yield the functions plotted in the logarithmic coördinates of Fig. 5.

Volume decreases as frequency increases, but the relation is clearly not a power-function. It also appears that volume is not a simple function of subjective pitch measured in mels. Of course, if frequency makes itself felt in the auditory system by reason of its ability to excite an anatomical location, it is not surprising that the subjective variables show no simple mathematical dependence on frequency.

Intensity seems to be another matter entirely. Volume grows with intensity according to the same basic psychophysical law—the power-function—that governs the growth of sensation on more than two dozen perceptual continua. With volume seeming to be another instance of this law, the record remains perfect: no prothetic continuum has failed to obey the power law. One wonders what will happen when tonal density is examined by these methods.[10]

A prothetic continuum is defined as one on which the variability, measured in subjective units, grows larger as the subjective magnitude increases. On a metathetic continuum the variability, in subjective units, tends to be constant. The evidence that volume behaves as a prothetic continuum is suggestive but not necessarily secure.

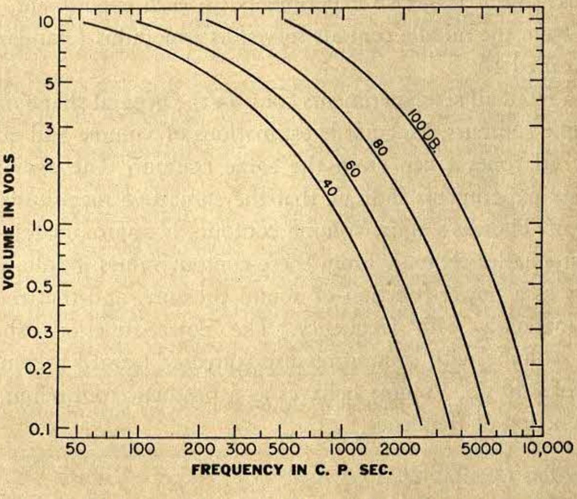


FIG. 5. RELATION BETWEEN VOLUME AND FREQUENCY
The curves represent horizontal cuts through the contours of Fig. 4.

Efforts to measure resolving power on this attribute—the volumic limen—have not been very successful.[11] It is significant, however, that the standard deviations that Thomas reported for an extensive series of equal-volume matches were approximately constant in logarithmic units (decibels).[12] Since the same was true for his equal-loudness matches, it appears likely that both volume and loudness behave as prothetic continua.

---

[10] Stevens, Tonal density, *J. exp. Psychol.*, 17, 1934, 585-592.
[11] Stevens, this JOURNAL, 46, 1934, 397-408.

SUMMARY

Tonal volume is defined as the apparent size or space-filling attribute of a sound. In six experiments groups of 10, *O*s estimated the magnitude of the apparent volume of tones selected from the equal-volume contours previously determined by Thomas. For each experiment two cuts, widely spaced, were made across each of five contours. These cuts defined a set of 10 tones, each of a different frequency and intensity. To check the reliability of the magnitude-estimations of volume, the tones from each cut were used in at least two experiments. In each experiment, one of the two tones from the middle contour served as a modulus (standard), whose value was called 10.

The data from all six experiments confirm the general shape of Thomas's equal-volume contours. Magnitude estimations of volume had quite similar values for all tones taken from the same contour. The averaged values from the six experiments indicate that the subjective magnitude associated with each of Thomas's equal-volume contours is approximately twice that of the next smaller contour. From these contour-values it follows that volume grows as a power-function of sound pressure, and that the exponent increases markedly with frequency. The power-functions showing the growth of volume with sound-pressure converge toward a common value at a level of 140 db. Volume behaves as a prothetic continuum and is not a power-function of frequency.

---

[12] Thomas, this JOURNAL, 65, 1952, 588-593.

THE EFFECT OF 'MEANING' ON THE AUTOKINETIC ILLUSION

By HANS H. TOCH, Michigan State University

The autokinetic illusion is conventionally described as "illusory movement presented by a small light of weak intensity in a dark room."[1] This definition is somewhat misleading, because the basic experience requires neither a weak light (pin-point source) nor a dark room. In fact, the first recorded observation of autokinesis, that of Alexander van Humboldt in 1799, took place on a mountaintop at daybreak. Humboldt noted that stars sometimes seem to perform oscillatory movements (*Sternschwanken*). Humboldt's astronomical observation was shown to be an illusion by Schweizer, who not only demonstrated that the stars 'oscillate' differently for different observers, but also duplicated the phenomenon in his laboratory. One of his demonstrations followed the conventional pattern of today (an opaque lantern, with a small opening, in a darkened room); another consisted of a black speck on a white background, which he reported to have behaved like an animated insect.[2]

The variety of sources which can produce autokinesis make it plausible to define the illusion in relatively general terms. Adams provides such a definition when he describes autokinetic movement as "illusory movement of a small object, or a collection of small objects, which is seen after a longer or shorter fixation against a relatively uniform background."[3]

Adams observes that when geometrical figures serve as stimuli for autokinesis, the nature of the movement is related to the shape of the stimulus. A 'round' light, for instance, "often gave the impression of a balloon, floating freely through the air"; it tended to turn in arc-like fashion, and to spin like a billiard ball; also, "it was possible for this light to go as fast in one direction as in any other, there being nothing in the shape of the light which would check movement in any direction." A 'square' light "often suggested a boat going through the water"; it tended to turn in sharp angles, and "like the round light, it could go in all directions with equal facility."[4] By contrast, a 'perpendicular, rectangular' light "moved much more frequently than did the square one to the right and left, less frequently up and

---

* Received for publication July 1, 1961. The experiments reported in this paper were supported by Research Grant G 8610 from the National Science Foundation. The author is indebted to Jane Ranney, who not only processed most of the subjects, but was also involved in the design of the study.

[1] H. A. Carr, The autokinetic sensation, *Psychol. Rev.*, 17, 1910, 42-75.

[2] S. Exner, Ueber autokinetische Empfindungen, *Z. Psychol.*, 12, 1896, 315.

[3] H. F. Adams, Autokinetic sensations, *Psychol. Monogr.*, 14, 1912 (No. 59), 1-32.

[4] Adams, *op. cit.*, 13.

down and along the diagonals."[5] The perpendicular light also showed differences in apparent speed. Whereas it tended to move slowly to the right or left, it moved fast upward or downward.

These observations with geometrical shape suggest the possibility that meaningful connotations in the stimulus-figure may also be able to determine the direction and quality of autokinetic movement. If a round light, giving "the impression of a balloon," tends to move in any of several directions, might not a figure shaped like an "airplane" tend to move in the direction of its "nose"? If a perpendicular rectangle tends to move left or right, could a perpendicular arrow not tend to move up or down, depending on where it was pointed?

Meaningful stimulus-figures have been used with some success to determine the direction of movement with various types of apparent movement, such as movements of shadows and after-images, induced movement, and stroboscopic movement.[6] The autokinetic illusion, which in its pure form provides no directional information, should lend itself even better than other forms of apparent movement to determination through 'meaning.'

Autokinesis has been experimentally modified by means of various types of experiences. The movement has been 'conditioned' for instance, by means of motor tasks and rewards and punishments.[7] The direction of the perceived movement has been determined by means of suggestive instructions.[8] Most important, individual *O*s have been made to yield their perceptual preferences in the face of pressures toward social conformity. They have come to perceive the autokinetic illusion in line with what they felt was being perceived by others.[9]

The use of 'meaningful' figures parallels the administration of social suggestion and the use of training techniques. In the latter cases, *O*'s perception is structured by experiences brought to bear on him before he perceives. In the case of 'meaningful' figures, past experiences are presumably built into the stimulus. Insofar as a figure is meaningful, it resembles objects which the subject has encountered in the past. The fact that the connotations of objects can play a role in structuring perceptual experience may be deduced from differences between the perception of 'meaningful' figures and their geometrical equivalents, or from differences among perceptions engendered through the use of differently 'meaningful' figures.

The present study explores the possibility of determining the direction of autokinetic movement through the use of figures representing objects with movement connotations. If perceived movement should tend to occur predominantly in 'meaningful' directions, the *O* has presumably responded in line with his past experience with relevant objects.

---

[5] *Ibid.*, 16.

[6] H. H. Toch and W. H. Ittelson, The role of past experience in apparent movement: A revaluation, *Brit. J. Psychol.*, 47, 1956, 195-207.

[7] E. A. Haggard and Rachel Babin, On the problem of "reinforcement" in conditioning the autokinetic phenomenon, *J. exp. Psychol.*, 38, 1948, 511-525; E. A. Haggard and G. J. Rose, Some effects of mental set and active participation in the conditioning of the autokinetic phenomenon, 34, 1944, 45-49.

[8] Muzafer Sherif, A study of some social factors in perception, *Arch. Psychol.*, 27, 1935, (No. 187), 1-60.

[9] Sherif, *op. cit.*, 1-60 also, *The Psychology of Social Norms*, 1936.

Apparatus and stimulus material. The apparatus used for the study was a portion of a larger set of equipment described elsewhere.[10] Illumination is provided by two 7½-w. filament bulbs contained in a 2½ x 4½-in. compartment fronted with diffusion glass. Between this diffusion glass and a screen of silvered glass facing the subject, a photographic transparency can be centered by means of a slideholder. For the present study, the light was switched on and off manually. Intensity was maintained constant (at approximately 11-ft. candles) with the aid of a voltmeter.

O sat 7 ft. directly in front of the screen, with his head immobilized in a headrest. In one set of conditions, he viewed the stimulus in a 5 x 5-in. square produced by covering the sides of a 5 x 15-in screen with black construction board. Another set of conditions faced *O* with a circular opening, 3½ in. in diameter, in a large round wooden disk, 30 in. in diameter. Observations took place in the dark.

The slides containing the stimulus-figures were opaque, with the stimulus-figure transparent, and all (except the control stimulus) approximately ¾ in. long. The stimuli were solid drawings, each occupying approximately the same area. Six figures were used: two (depicting a running man and a jumping deer, respectively), were employed to elicit horizontal movement; two (an airplane and a bomb or rocket) were used for vertical movement; one (an arrow) was used either horizontally or vertically; and another (a small solid disk) was used as a control stimulus.

General procedure. Each *O* was presented with a series of four figures, the control stimulus (small disk) and three 'meaningful' figures. The three figures either fell into the horizontal or the vertical group. They were alternately pointed in opposite directions. The specific figures were systematically randomized, and the *O*s were randomly assigned to a given sequence. Typical sequences, for instance, were (1) disk; running man, facing left; jumping deer, facing right; and arrow, pointed left; (2) disk; arrow, pointed down; airplane, pointed up, and bomb (rocket), pointed down.

The *O*s were instructed as follows:

> I shall show you a light. After a while, this light will start to move. Tell me *at once,* as soon as the light starts to move. Please also tell me in what direction the light is moving.

PILOT STUDY

In the pilot study, 26 *O*s were presented with the vertical figures (airplane; bomb or rocket; arrow) in the square screen. The data obtained in this experiment are summarized in Table I. As may be noted, three out of five of the movements perceived with each figure, proceeded in the 'correct' direction (up or down). For each of the test-figures, a Chi-square test of independence (corrected for continuity) was computed to see whether the number of persons perceiving movement in the 'correct' and 'incorrect' vertical direction was independent, *i.e.* whether it departed significantly from 50%. In the case of the airplane and the arrow, the chi-squares were significant ($p < 0.02$ and $p < 0.01$ respectively). The difference in the case of the bomb or rocket fell short of significance ($p < 0.10$).

---

[10] Toch and Ittelson, *op. cit.,* 203 f.

Main Study

In the main study, 150 *O*s were exposed to the autokinetic illusion using the 'horizontally meaningful' figures (arrow, running man, and jumping deer). The square screen was used with roughly half this group (73 *O*s).[11] The data obtained show an anomalous distribution of movements for the control figure, which seemed to be related to the shape of the screen. The circular screen was constructed to control for this possibility, and the remaining 77 *O*s were processed with it. Table II shows the 'control' observations of *O*s exposed to the square and round screens. With the square screen, two thirds of the perceived movements occurred on the horizontal plane—over three times as many horizontal movements as vertical movements. This horizontal preference is not evident with the circular screen. A horizontal-vertical Chi-square test of independence proved signficant for the square screen ($p < 0.001$), but not for the circular frame. The circular frame also frequently elicited definite circular and ambient (three dimensional) movements, which only rarely occurred in passing using the square screen.

TABLE I

Autokinetic Movements Observed with 'Meaningful' Stimulus-Figures that Suggest Vertical Movement

(N=26)

Stimulus-figure (Pointed up or downward)	Horizontal movement			Vertical movement		Other kinds
	none	left	right	correct	incorrect	diagonal
Airplane	1	0	5	16	4	0
Arrow	0	2	2	16	3	3
Bomb	0	1	3	15	6	1
Total	1	3	10	47	13	4
Control (circle)	0	3	7	9*		7

* Four movements upward and five downward.

Table III presents the data for the 'meaningful' figures, as viewed by the 150 *O*s.[12] The Chi-square test of independence was applied to the horizontal movement of the arrow, and to the horizontal and diagonal movements of the running man and jumping deer (which were somewhat ambiguous as to plane of movement). The difference in favor of 'correct' movement proved significant for each figure ($p < 0.001$).

---

[11] Seventeen additional *S*s were used with the square screen and the control stimulus.

DISCUSSION

Our data show that meaningful connotations in an autokinetic stimulus participate in determining the direction of the perceived movement. A figure depicting an object which has been experienced as moving seems to include directionality among its connotations. This directional connotation may facilitate the perception of movement in the implied direction, and reduce the probability of movement in other directions.

TABLE II

EFFECT OF FRAME ON THE PERCEPTION OF MOVEMENT WITH A DISK FIGURE

		Horizontal movement				Vertical movement			Other kinds	
Frame	*N*	none	left	right	both	up	down	both	diagonal	circular or ambient
Horizontal	90	0	26	33	1	9	9	1	11	0
Circular	77	1	21	12	—	13	6	—	12	12

The autokinetic illusion is probably the most sensitive available test for determinants of the direction of movement, since the stimulus itself is structurally neutral. It occasions movement *per se,* and seems to leave the direction, extent, speed, and duration to *O*. Any subjective and structural biases in the perceptual transaction can function to supplement the inconclusiveness of the stimulus. Fixation, 'will power,' and 'suggestion' have been cited among such biasing factors.[13] Eye dominance or other functional non-equivalence along the optic pathways could act as biasing determinants. The experiment reported above shows that 'meaning' can also play this role.

Admittedly the nature of autokinetic movement is such that relatively 'weak' determinants can enter to determine what is perceived. All that can be inferred about the effectiveness of any determinant in this type of perceptual situation is that it *can* function as a determinant. Thus we can conclude no more from the above than that meaningful connotations *can* influence movement direction—at least for *some* *O*s, under *optimal* conditions.

---

[12] Pooling the data for the two viewing conditions seemed justified because (a) both screens are symmetrical, so that no left or right bias would be expected; and (b) each stimulus was pointed left or right 50% of the time, so that *even if there was* a directional effect, it could not qualify the results.

[13] Exner, *op. cit.,* 313, 330; Carr, *op. cit.,* 75; Adams, *op. cit.,* 32; Sherif, *op. cit.,* 60.

It is also not legitimate to conclude from our data that the autokinetic phenomenon is *centrally* determined. To be sure, it is not likely that 'meaning' can operate at the retinal or oculo-motor level directly. We may not conclude, however, from this that autokinesis does not depend for its effectiveness on peripheral determinants which occasion apparent movement nor that retinal mechanisms are not used in the service of meaning. It is very likely, in fact, that the determining pattern is so complex, and the causal relationships so mutually dependent, that one can not speak of a 'sequence' of determination, except didactically. The central-peripheral controversy, here as elsewhere, may be an artifact, analogous to selecting the 'winner' in a horse race where the animals run in all directions simultaneously.

TABLE III

AUTOKINETIC MOVEMENTS OBSERVED WITH 'MEANINGFUL' STIMULUS-FIGURES THAT SUGGEST HORIZONTAL MOVEMENT

(N=150)

Figure (pointed left or right)	none	Vertical movement			Horizontal movement			Other kinds		
		up	down	both	correct	incorrect	both	right diagonal	left diagonal	circular or ambient
Arrow	9	12	8	2	69	25	5	0	3	17
Running man	8	15	11	1	63	15	6	13	4	14
Jumping deer	9	14	8	3	47	20	4	21	10	14
Total	25	41	27	6	178	60	15	33	17	45

Among the determinants of autokinetic movement no mention seems to have been made of the shape of a 'frame.' This is understandable since, in theory, autokinesis presupposes complete darkness or a homogeneous field. In practice, when a visible frame has been used, the shape appears to have been a constant.[14] Our data indicate that a physical frame can become a psychological frame of reference for autokinesis. In the case of our partially covered rectangular screen, movement appears to have been channelled into the horizontal plane. To what extent is this effect generalizable? How would differently shaped frames function? Subsequent experimentation should answer these questions.

---

[14] E. L. Hoffman, D. V. Swander, S. H. Baron, and J. H. Rohrer, Generalization and exposure-time as related to autokinetic movement, *J. exp. Psychol.*, 46, 1953, 171-177.

SUMMARY

The study explored the effectiveness of meaning connotations in determining the direction of autokinetic movement. 'Meaningful' figures were presented on a screen in a dark room. A preliminary experiment used stimuli pointed upward or downward, and the main experiment used figures pointed left or right. In both instances, the autokinetic movement was found to occur to a large extent in the 'meaningful' direction. The use of two differently shaped screens also showed that the direction of autokinetic movement can vary with the shape of the screen. The implications of these findings were discussed.

SINGLE-STIMULUS JUDGMENTS OF LOUDNESS AS A FUNCTION OF PRESENTATION-INTERVAL

By WILLIAM BEVAN and JOAN F. PRITCHARD, Kansas State University, and WALTER G. REED, Ohio State University

It is a common observation that when two stimuli of equal intensity are presented in succession in one of the comparison methods, the magnitude of the second is judged to differ from that of the first with greater frequency than is attributable to chance. This is the time-order error. It occurs often in loudness- and heaviness-judgments and usually is measured as the difference between the Indifference Point (*IP*) of the series and the Standard (the middlemost) Stimulus.[1] The time-order error (*TOE*) relates to a number of variables, preëminent among which are the temporal interval between successive stimuli,[2] the general level of stimulation,[3] and the psychophysical method employed.[4]

The categorical methods yield a similar effect. The *IP* generally differs from the middle of the stimulus-series. This is the series-effect. Both series effects and *TOE*s have been effectively dealt with an adaptation-level (*AL*) theory.[5] The relation of the series-effect to level of stimulation was early recognized as the Central Tendency Effect,[6] and the *IP* as a function of psychophysical method has received systematic attention.[7] Meanwhile, although Philip has expanded the basic *AL* formula to include the time-interval between stimuluation and response in the method of reproduc-

---

* Received for publication July 7, 1961. This experiment was begun at Emory University and completed at Kansas State University. The work at Kansas State University was supported by the Faculty Research Fund and by the Office of Naval Research, the latter through contract Nonr-3290(01), Project NR 142-155. The writers are indebted to Mrs. Barbara Maier for her aid in testing at Emory University.

[1] J. P. Guilford, *Psychometric Methods*, 2nd ed., 1954, 306.

[2] Wolfgang Köhler, Zur Theorie des Sukzessivvergleichs und der Zeitfehler, *Psychol. Forsch.*, 4, 1923, 115-175.

[3] J. G. Needham, Contrast effects in judgments of auditory intensities, *J. exp. Psychol.*, 18, 1935, 214-226.

[4] Leo Postman, Time-error as a function of method of experimentation, this JOURNAL, 60, 1947, 101-108.

[5] Harry Helson, Adaptation-level theory, in Sigmund Koch, *Psychology: A Study of a Science*, 1, 1959, 565-622; W. C. Michels and Harry Helson, A quantitative theory of time-order effects, this JOURNAL, 67, 1954, 327-334.

[6] H. L. Hollingworth, The central tendency of judgment, *J. Philos. Psychol. sci. Meth.*, 1, 1910, 461-469.

[7] Helson, Adaptation-level as frame of reference for prediction of psychophysical data, this JOURNAL, 60, 1947, 1-29.

tion,[8] no attention has been directed toward the empirical investigation of the series-effect and presentation-interval. Accordingly, the purpose of the present experiment is threefold: (1) to determine the influence of presentation-interval upon the *IP* with the single-stimulus-method; (2) to examine this relationship over a wide range of intervals (20 sec. appears to be the longest interval used to date with the comparison-methods)[9]; and (3) to explore the relation of presentation-interval to the effectiveness of anchors.

(1) Design. This study consists of two experiments: one involving a soft series of tones to be judged for loudness, with an anchor above the series; and the other a loud series, with the anchor below. Each experiment employed 12 groups of *S*s, a control (no anchor) and an experimental (anchor) group for each of 6 presentation-intervals. The stimulus-series in both experiments consisted of 5 intensities of a 1000~ tone 1 sec. in duration. The soft series: 30, 35, 40, 45, and 50 db. (*SPL*); the loud series: 55, 60, 65, 70, and 75 db. (*SPL*). The anchor above the soft series was 70 db.; that below the loud series was 35 db. The presentation-intervals selected for use were 2, 4, 8, 16, 32, and 64 sec.

(2) Subjects and sessions. A total of 240 *S*s, all women, served in the study: 10 in each of the 12 groups of the 2 experiments. The *S*s receiving the 'soft' series were drawn from a recent summer session of the George Peabody College for Teachers; those receiving the 'loud' series were recruited from the introductory course in Psychology at Kansas State University. None of the *S*s was an experienced psychophysical judge.

Each *S* was tested on two occasions separated by about one week. Members of the control groups made 70 judgments—20 practice and 50 test—with the series-stimuli during each session. Each series-stimulus occurred with the same frequency. Members of the experimental groups made the same number of series-judgments as did the control *S*s. These were supplemented by presentation of the anchor-stimulus as every fifth trial.

(3) Apparatus. The 1000~ tone was generated for the 'soft-series' group by a Jackson Audio-oscillator, Model 655, for the 'loud-series' groups by a Hewlett-Packard wide range oscillator, Model 200CD. Stimulus-intensity was controlled in both experiments by means of a Hewlett-Packard 5-w. 600-Ω attenuator, Model 350B. A Wichita electronic switch was used for the soft-series, and a Grason-Stadler electronic switch, with the loud to control rise-time and to eliminate clicks. The stimulus-tones were presented to the soft-series groups through Permaflux Model P.D.R.-8 earphones and to the loud-series groups through Telephonics Model TDH-39 phones. The duration of both stimulus- and presentation-intervals in the two experiments was controlled by means of two Hunter Model 111C silent relay-interval-timers so wired as to recycle automatically.

(4) Procedure. The method used was the rating-scale variation of the method

---

[8] B. R. Philip, The frame of reference concept, *Canad. J. Psychol.*, 3, 1949, 73-79.

[9] E. König, Effect of time on pitch discrimination thresholds under several psychophysical procedures: Comparison with intensity discrimination thresholds, *J. acoust. Soc. Amer.*, 29, 1957, 606-613.

of single stimuli. At the beginning of each session *S* was instructed that she was to listen only for the tones presented through the headphones and to disregard extraneous noises. She was to make her judgments with reference only to the tones she was to be presented during each particular session. A 13-category scale was placed before her and it was explained that when a tone appeared neither loud nor soft but medium it was to be called zero. When it appeared louder than medium, its intensity was to be described, following the procedure of Webb and Bevan, as some appropriate multiple of 100 above zero;[10] when it appeared softer, it was to be rated with some appropriate multiple of −100. Finally, it was explained that the scale to be used was open-ended at both extremes and that the judge could use whatever value, positive or negative, she deemed accurate to express her judgment.

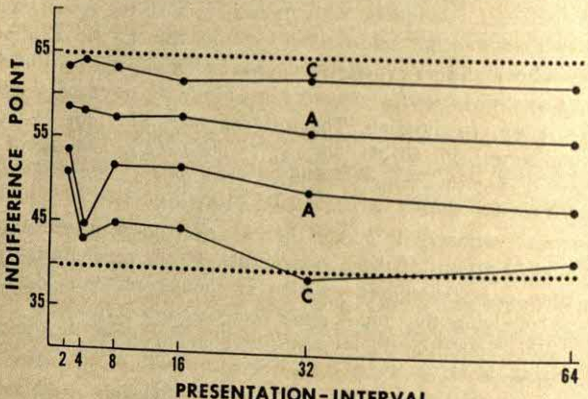


FIG. 1. INDIFFERENCE POINTS FOR LOUDNESS AS A FUNCTION OF LEVEL OF STIMULATION, PRESENTATION-INTERVAL, AND THE PRESENCE OF AN ANCHOR IN THE JUDGMENTAL SEQUENCE

The dotted lines represent the midpoints of two arithmetic series of auditory intensities. Curves marked 'C' are for groups presented the series alone; those marked 'A' indicate the influence of an anchor above and below the series, respectively. Every point represents the mean IP of 10 *S*.

The median magnitude judged for each series-member and for the anchor were obtained for each individual *S*. Each *S*'s *IP* (median judgment) was then calculated by linear interpolation. The group *IP*s represented in Fig. 1 are means of these individual values. An Alexander trend-test was performed on the *IP*s for each experiment.[11]

RESULTS AND DISCUSSION

The series-effect and general stimulation-level. The dotted lines in Fig. 1 indicate the midpoints of the two arithmetic stimulus-series. It will be noted that the series-effect in the soft-series control is positive, the *IPs* being in all instances except one above the midpoint. The exception, at 32 sec., lies so close to the midpoint as to be of doubtful negativity. In contrast, all control groups in the intensive series gave negative *IPs*. This is a clear example of the central-tendency phenomenon. It would appear to resemble descriptions of the relation of loudness *TOEs* to stimulus-intensity. Needham reports predominantly positive *TOEs* for 30- and 40-db. tones with predominantly negative *TOEs* for 50- and 60-db. stimuli,[12] and Koester claims medium intensity to be associated with positive and loud stimuli with negative errors.[13]

---

[10] S. C. Webb and William Bevan, Single standard judgment: A reorientation in rating scale procedure, *J. gen. Psychol.*, 59, 1958, 119-133.

[11] H. W. Alexander, A general test for trend, *Psychol. Bull.*, 43, 1946, 533-557.

The series-effect and presentation-interval. The *IPs* for both series are seen to drop with an increase in time-interval. While the slope of the control curve for the loud series is relatively gentle and appears to be linear (over-all deviation from linearity and group-deviations from estimate Ps both > 0.05) and the curve for the soft series is more precipitous and negatively accelerated (over-all deviation from linearity, $P > 0.01$; group deviations from estimate, $P > 0.05$), the curves for *both* series display significant negative over-all slope (over-all slope, soft series, $P < 0.01$; loud series, $P < 0.05$). These data appear to contrast with *TOE*-data in several other respects. Köhler, in a frequently cited study, described the loudness *TOE* as positive for short intervals (1–3 sec.) and negative for longer (4–12 sec.).[14] This was corroborated for the constant stimulus method by Postman, who, however, failed to note any trend for the method of average error.[15] Koester could find no systematic relationship between *TOE* and presentation-intervals up to 9 sec.[16] Needham, meanwhile, has shown the relationship to be a complex one, involving an interaction between presentation-interval and stimulus-intensity:[17] the *TOE* for low intensities becomes increasingly positive with increased intervals up to 6 sec. The opposite is true for moderately intense sounds. Further, the magnitude of the series-effect is large compared to the magnitude of the typical *TOE*. Stevens describes the latter as small—usually a fraction of a *jnd*.[18] The series-effects observed in the control data vary between 1.4 and 10.3 db.

---

[12] J. G. Needham, The effect of the time interval upon the time error at different intensive levels, *J. exp. Psychol.*, 18, 1935, 530-543.

[13] Theodore Koester, Time error and sensitivity in pitch and loudness discrimination as a function of time interval and stimulus level, *Arch. Psychol.*, 32, 1945 (No. 297), 1-69.

[14] Köhler, *op. cit.*, 151.

[15] Postman, *op. cit.*, 103-104.

[16] Koester, *op. cit.*, 59-60.

[17] Needham, 1935, *op. cit.*, 534-536.

[18] S. S. Stevens, On the psychophysical law, *Psychol. Rev.*, 64, 1957, 153-181.

and between 0.7 and 3 db. respectively for the soft and loud series. Finally, unlike the relationship between presentation-interval and *TOE* which may change radically with repeated testing,[19] the slope of the present curves remained the same for the first and second sessions a week apart.

The consistently negative slope of the present curves graphically reflects the common observation that contrast effects are enhanced by prolonged prior stimulation of contrasting intensity. It is easily reconciled with the principle of pooling identified with *AL* theory.[20] Several assumptions must, however, be made explicit. It must first be assumed that the organism integrates input at a more or less constant rate for given circumstances and that this rate is the same within the several conditions of the present experiments. Secondly, it must be assumed, with Lauenstein, that silent periods are integrated as zero inputs.[21] It thus follows that as presentation-interval increases the number of zeroes integrated correspondingly increases and the *AL* (*IP*) is lowered. It is to be expected that, depending on the rate of integration and the length of the presentation-sequence, the *IP* will sooner or later reach some minimal value for the stimulus-series under consideration and will not change with further increases in presentation-interval. The present data suggest that these values were reached with intervals of approximately 32 sec. for the soft series and 16 sec. for the loud. Parducci has recently reported an empirical study of the recency-effect in single-stimulus judgments.[22]

Anchor-effects and presentation-interval. Just as the pooling model explains the dependence of *IP* upon presentation-interval, it also suggests that the effectiveness of anchors is influenced by the same variable. If, as presentation-interval increases, the number of zero inputs also increases and *AL* is lowered, then the scalar interval between *AL* and anchor-magnitude should increase, and the effect of the anchor be enhanced. In the case of the soft series, this means that the slope of the curve for the anchor-groups should be less than that of the control groups. In contrast, in the case of the loud series, the slope of the anchor-curve should be greater than that of the control groups. When smooth curves were fitted to the data of Fig. 1, they suggested confirmation of these hypotheses. Statistical analysis bore this out, however, only in the case of the soft series (soft series: between control and anchor-group slopes $P < 0.05$; loud series: between control and anchor-group slopes $P > 0.05$). The reason for the failure of confirmation for the loud series is a matter for conjecture. One possibility concerns the particular spacing of series and anchors used in these experiments. It has been frequently reported—and this is confirmed in Fig. 1—that anchors above a series are more effective than anchors below.[23] In the present case, it took much preliminary experimentation before a combination of loud series and soft anchor could be found that would yield a reliable anchor effect. It is conceivable that in selecting a soft anchor that would be effective, the experimenters selected one that masked the differential effect of presentation-interval.

---

[19] J. G. Needham, Time error in comparison judgments, *Psychol. Bull.*, 31, 1934, 229-243.

[20] Helson, Adaptation-level as a basis for a quantitative theory of frames of reference, *Psychol. Rev.*, 55, 1948, 297-313.

[21] Otto Lauenstein, Ansatz zu einer physiologischen Theorie des Vergleichs und der Zeitfehler, *Psychol. Forsch.*, 17, 1933, 130-177.

[22] Allen Parducci, Learning variables in the judgment of single stimuli, *J. exp. Psychol.*, 48, 1954, 24-30.

The dip between 2–8 sec. in the curves of the soft series is at present unclear. Since this is not duplicated in the curve for the loud series, one may possibly attribute it to experimental artifact. At the same time, the fact that it occurs for both conditions of the soft series raises a question about this conclusion. One possibility to be considered involves qualitative differences (rapid reaction versus reflective judgment) in the response-mechanisms associated with short and long intervals, respectively.

An attempt to account for any relationship between series-effects, stimulation-level, and presentation-interval and anchors in other than judgmental terms is at present premature. Galambos and Rupert have recently reported that contractions of the muscles of the middle ear of the cat are capable of reducing transmission of loud tones (60 db. or more) across the middle ear by as much as 20 db.[24] This could account for the positive series-effect in the soft series and its attenuation with increased presentation-intervals. It would also explain the elevation of the *IP*s by the anchor. The data of the loud-series is not so readily accommodated. Meanwhile, if it may be assumed that response of the muscles of the middle ear reflects changes in relative rather than absolute intensity—in vision, for example, accommodation and convergence have been related to apparent in contrast to physical distance—then the negative series-effect and the anchor-effect of the soft-series can more readily be explained.[25]

Judgments of individual S*s.* Perhaps the source of variance most frequently met in psychological experiments is that associated with differences between subjects. A striking characteristic of the present data is that the F-ratios for between individual slopes and between individual means in both experiments are less than 1.00. This apparent absence of individual differences could only result from homogeneous groups making a large number of judgments on a relatively simple judgmental task.

---

[23] Cf. Helson, 1948, *op. cit.*, 309-310.

[24] Robert Galambos and Allen Rupert, Action of the middle ear muscles in normal cats, *J. acoust. Soc. Amer.*, 1959, 31, 349-355.

[25] W. H. Ittelson and Adelbert Ames, Jr., Accommodation, convergence, and their relation to apparent distance, *J. Psychol.*, 1950, 30, 43-67; S. Ohwaki, On the factors determining accommodation: Research on size constancy phenomena, *Tohoku Psychol. Folia,* 14, 1955, 147-158.

SUMMARY

Two hundred forty *S*s (women) made a total of more than 40,000 single-stimulus judgments of loudness. Half of them judged a 'soft' series and half a 'loud' series; and half received an anchor above or below the series and half judged the series alone. Successive presentations for individual *S*s were at one of six time-intervals: 2, 4, 8, 16, 32, or 64 sec.

Series-effects for the soft series, with one slight exception, were all positive. Those for the loud series were negative, Anchors, when added to both series, significantly enhanced the series-effect. All curves relating presentation-interval to loudness *IP*s display a reliable negative slope. This is explained in terms of the pooling model of adaptation-level theory. An anchor by presentation-interval interaction, predicted from the same theory, was observed for only the soft series.

APPARENT INTERPOSITION IN BINOCULAR DEPTH-PERCEPTION

By RODERICK P. McDONALD, University of New England, Australia

In the normal environment, the major stimulus-variables governing the perception of depth are commonly so closely associated that special experimental techniques must be developed to manipulate them independently. Thus Carr's stereoscope was used by Swenson to dissociate the stimulus-variables involved in accommodation and binocular disparity.[1] Similarly, Schriever employed the stereoscope in an investigation of interposition and binocular disparity in which these 'cues' were set in conflict, as it is commonly put, by presenting the stereo-pairs pseudoscopically.[2] Schriever's investigation is widely quoted, for example by Koffka, Woodworth and Schlosberg, and Brunswik, as evidence for a quite general assertion that in such a conflict the stimulus-information from binocular disparity will cease to be an effective determinant of the judgment.[3]

Schriever's stereoscopic method is virtually limited to the study of judgments of relative depth, hence there seemed to be a case for testing the generality of this assertion under conditions that would permit other types of psychophysical measurement, such as matching the distance of an adjustable marker to the apparently masking and the apparently masked object. The problem set was to arrange for the apparent masking of one object by another that is physically behind it, much as in Ames' well known demonstration but so designed as to be equally effective in binocular vision.[4] The solution essentially consisted in the use of a real image formed in space by an optical system as the object to be masked.

METHOD

Apparatus. A horizontal section of the apparatus is shown in Fig. 1. A disk of paper, *Do*, about 1 in. in diameter was pasted over a microscope-illuminator *I*,

---

* Received for publication February 15, 1961. The author is indebted to Shirley Barratt for her assistance in the investigation.

[1] H. A. Swenson, The relative influence of accommodation and convergence in the judgment of a distance, *J. gen. Psychol.*, 7, 1932, 260-380.

[2] Walter Schriever, Experimentelle Studien über stereoskopisches Sehen, *Z. Psychol.*, 96, 1925, 113-170.

[3] Kurt Koffka, *Principles of Gestalt Psychology*, 1935, 274; R. S. Woodworth and Harold Schlosberg, *Experimental Psychology*, 1954, 462-463; Egon Brunswik, *Perception and the Representative Design of Experiments*, 1956, 51.

[4] C. E. Osgood, *Method and Theory in Experimental Psychology*, 1953, 263-265.

and placed about 1 ft. away from the convex face of a large plano-convex lens *L* (6 dioptres, 11 cm. in diameter). A real image *Di* of the disk was formed about 4 in. in front of the convex face of the lens by reflection from the planosurface and focusing at the convex surface. The apparatus was enclosed in a light-tight box with provision for seating *O* at one end and *E* at the other, as shown. When viewed binocularly through the two slits, S_1 and S_2, (each 4 in. long and 1 in. high) the image of the disk appeared in the air-space as a visually hard surface with a quite definite spatial localization. There were no detectable optical aberrations in the image. A virtual image and a further real image could be seen behind the disk *Di* from *O*'s position. These were eliminated with masking tape applied to the surface of the lens.[5] The distance from the outer slit to *Di* was approximately 75 cm. A masking card *C* was hung on a wire-track across the box about 2.5 cm. behind the disk *Di* as seen by *O*, with stops on the track so that *E* could reach into the box and move the card to one of two extreme positions. In the 'mask in' (*MI*) position, the card intercepted part of the image-forming flux of light coming back from *L*, so that from *O*'s point of view the card appeared to overlap the disk *Di*, covering about a quadrant of it. In the 'mask out' (*MO*) position, from *O*'s point of view the card was not quite overlapping the disk. A pointer *P* was so mounted on a meter-stick *M* that it could move along a line parallel to *O*'s line of sight. Adjustments of the pointer could be made by direct manipulation of the end of the meter-stick, which projected from the screen carrying the slit S_2. The screen carrying the slit S_1 extended only part of the way down from the top of the box. At the other end of the box, *E* could read off the settings of the pointer in millimeters from the other projecting end of the meter-stick where it contacted a fixed marker, and could move the masking card on its track as required, with access to the interior of the box via a set of black curtains.

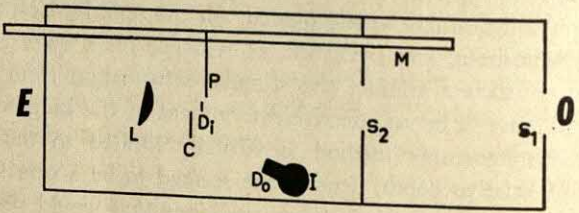


FIG. 1. PLAN OF THE APPARATUS

The angular subtense of *Di* at *O*'s position was approximately 90′. The angular separation of the tip of the pointer and the visible edge of the card was 60′ in Condition *MI* and 100′ in Condition *MO*. Other things equal, the increased separation in Condition *MO* might be expected to lead to poorer stereoacuity in this condition, but should not produce any systematic effects on either the distance-matching observations or on relative judgments of depth. Any artifact of this kind in the data on stereoacuity can be tested for in the analysis.

---

[5] In principle, a concave mirror could be used in place of the lens. It was found by trial, however, that the masking out of extraneous images was too difficult when a mirror was employed.

Procedure. Twelve undergraduate students of psychology served as *O*s. The instructions given were as follows:

> When you look through the slit, you will see a circular disk, a corner of white card, and a pointer. I want you to adjust the pointer so that it looks alongside either the disk or the card, whichever I specify. You may move the pointer back and forth until you are satisfied with your adjustment.

Each *O* made 16 observations, consisting of 4 replications of each of 4 conditions. The 4 conditions consisted of matching the pointer either to the disk or the card in either Condition *MO* or Condition *MI*. To provide counterbalancing for the order in which the *O*s experienced the 4 conditions, the 12 *O*s were divided into 4 groups, the observations in each group being made in appropriate fixed sequences.

The mean of each *O*'s settings of the pointer under each condition provided data on distance-matching (points of subjective equality—*PSE*s), while intra-observer variability within a condition provided a measure of stereoacuity. In addition, at the end of the session each *O* was asked the neutrally phrased question: "Have you any comments on what you saw?" Without further prompting, it was possible to classify responses to this question so as to yield information about *O*'s direct impression of the relative distances of the disk and the card. Other aspects of these responses will also be discussed below.

RESULTS

Preliminary analysis of the distance-matches showed that there was no evidence of effects due to the order in which the judgments were made, so an analysis of variance was carried out on the means of the four *PSE*s given by each *O* under each condition, yielding a mixed model consisting of Disk vs. Card by *MI* vs. *MO* by *O*s. The main effect of Disk vs. Card gave an F of 19.563 on 1 and 11 df ($p < 0.001$). The interaction between Disk vs. Card and *MI* vs. *MO* gave an F of 2.266 on 1 and 11 df, which is not significant. In Condition *MI*, the mean setting for the card was 2.76 cm. behind the mean setting for the disk. In Condition *MO*, the mean difference was 2.14 cm. in the same direction. Hence in the present investigation the sucessive distance-matchings made to the apparently masked and the apparently masking object were unaffected by the apparent interposition.

All the *O*s reported great difficulty in making distance-matches at all in the masking condition (*MI*). A plausible hypothesis is that this subjectively experienced difficulty would be reflected in a wider scatter of settings in this condition. As a computationally convenient yet fully efficient measure of stereoacuity, for each *O* under each condition the absolute difference of the means of pairs of *PSE*s was taken. An analysis of variance parallel to the one for the *PSE*s yielded no significant effects, and in particular the value of F for the main effect of masking on stereoacuity was

0.746. The possible artifact mentioned above, from the variable separation of the edge of the card and the pointer would be expected to reveal itself in the interaction between Disk vs. Card and *MI* vs. *MO*. For this interaction, *F* was 0.968.

From *O*s' introspective reports it can be stated that 10 out of 12 were "not sure" which object was in front of which in Condition *MI*. The one *O* who was sure that the card was in front in this condition revealed poor stereoacuity in all conditions, and could reasonably have been eliminated from the group by a preliminary test of binocular functioning. The lack of certainty in the simultaneous depth-comparisons is in marked contrast with the precision revealed in the successive distance-matchings to the disk and the card separately.

Another effect demonstrated in the investigation can only be illustrated by a sample of statements from the protocols:

> "I couldn't focus the things together." "The disk seemed at the back, yet when I moved the pointer it seemed in front." "The disk seemed in front till I shut one eye and realized that it was really at the back." "The disk looked in front. I thought they were even, really, but they didn't look it, so I judged what I saw, not what I thought."

In effect, the statements reveal some sort of puzzlement as to how what is in front of them in the masking condition can have been brought about.

DISCUSSION

Under the conditions of the present investigation, distance-matchings of a pointer to the masked and the masking object were determined by binocular disparity, the conflicting interposition being quite ineffective. The difficulty reported by the *O*s in making these matches was not reflected in the matches themselves. On the other hand, it seems that the direct depth-comparisons of the masked and the masking object were influenced by the apparent interposition, though not to the extent of outright dominance of interposition as found by Schriever using the stereoscopic method. In effect, while *O* is making a distance-match, he can attend to the disk or card and to the pointer, and match them on the basis of binocular disparity. In a direct comparison of the distances of the disk and the card it is reasonable to expect a direct conflict between the stimulus-information from the interposition and that from the disparities.

Finally, it may be pointed out that in a situation such as the present one, we set up what Brunswik would call a conflict between two cues of high ecological validity.[6] Reactions to perceptual incongruity could thus be expected in terms of the classification given by Bruner and Postman; namely, dominance-reactions, compromise-reactions, disruption, and the recognition of incongruity.[7] Most of the protocols fall without ambiguity into the last category, as exemplified by the sample of statements given above. These last considerations are perhaps somewhat speculative, but it is interesting to note the applicability of Bruner and Postman's concepts to a depth-perception situation where the stimulus-conditions are not minimal.

---

[6] Brunswik, *op. cit.*, 50-51.

---

[7] J. S. Bruner and Leo Postman, On the perception of incongruity: A paradigm, *J. Personal.*, 18, 1949, 206-223.

PICTORIAL RECOGNITION AS AN UNLEARNED ABILITY: A STUDY OF ONE CHILD'S PERFORMANCE

By JULIAN HOCHBERG and VIRGINIA BROOKS, Cornell University

Anecdotes about primitive people who are unable to identify pictured objects suggest the hypothesis that pictorial recognition is a learned ability.[1] In a weaker form of this hypothesis, learning might be held essential for the recognition of line-drawings (compare Gibson's 'ghost shapes'),[2] while the naïve recognition of photographs, with their higher 'fidelity,' would be admitted. The present investigation was designed to determine whether a child who had been taught his vocabulary *solely* by the use of objects, and who had received no instruction or training whatsoever concerning pictorial meaning or content, could recognize objects portrayed by two-dimensional line-drawings and by photographs.

Answers to these questions were desired for two reasons. First, the psychophysical exploration of outline-representations has begun to provide some promising, lawful relationship.[3] Although the predictive equations would remain just as interesting regardless of whether they are based on 'learned' or 'innate' processes,[4] somewhat different sets of further hypotheses might suggest themselves if the entire realm of responses to outline-representations of spatial objects turned out to be the product of arbitrary associations between symbols and things—a sort of assigned visual language. Secondly, if pictorial perception did indeed turn out to be a learned ability in this arbitrary sense (which it did not), we should have a starting point for the investigation of the possible lines of its development and of individual differences therein.

It should be stressed that this investigation does not directly bear upon the general question of nativism vs. empiricism in space-perception, which is too broad to be submitted to so specific a test. If space-perception were itself an 'unlearned' ability, the representation by flat pictures might not be feasible without specific learning. If recognition of solid objects in two-dimensional representation were at least in part unlearned, it might still

---

* Received for publication April 18, 1961. This study, assisted by Dr. P. C. Smith and Mrs. Janice Goldstein, was supported by an NIH research grant, B-1586(C2).

[1] The only study in point is that of W. Hudson, Pictorial depth perception in sub-cultural groups in Africa, *J. soc. Psychol.*, 52, 1960, 183-208, which is concerned with differences in spatial localization rather than in recognition of objects.

[2] J. J. Gibson, What is a form?, *Psychol. Rev.*, 58, 1951, 403-412.

[3] Julian Hochberg and Virginia Brooks, The psychophysics of form: Reversible-perspective drawings of spatial objects, this JOURNAL, 73, 1960, 337-354.

[4] Hochberg, Spatial representation: Theme 10, *Proc. Int. Congr. Psychol.*, 1957.

develop without formal training as a by-product of a more general process of learning to perceive space.

Training. Since birth, the subject (*S*), a boy, had been exposed to and taught the names of a wide variety of toys and other solid objects. With two exceptions discussed below, the color of each of these objects was either uniform, or it was divided into functional areas (*e.g.* face-coloration, hair-shade, and dress-color on dolls). That is, no objects were depicted as surface-decoration. Even so, *S* never was told (or allowed to overhear) the name or meaning of any picture or depicted object. In fact, pictures were, in general, kept from his immediate vicinity.

TABLE I

STIMULUS-PRESENTATIONS AND JUDGES' INTERPRETATIONS OF *S*'S RESPONSES

Part 1				Part 2			
	Interpretation of responses				Interpretation of responses		
Stimulus-picture	Parents	Judge A	Judge B	Stimulus-picture	Parents	Judge A	Judge B
(1) car	+	car	car	(1a) car	+	car	car
(2) car	+	car	car	(2a) car	+	car	car
(3) shoe	+	shoe	shoe	(3a) car	−	Jody§	Jody§
(4) shoe	+	shoe	shoe	(4a) shoe	+	shoe	shoe
(5) shoe	+	——	——	(5a) shoe	−	——	——
(6) Jody (sister)	+	Jody	——	(6a) Jody	+	Jody	Jody
(7) dolly	+	——	dolly	(7a) dolly	+	dolly	dolly
(8) dolly	−	shoe‡	shoe‡	(8a) dolly	+	dolly	dolly
(9) dolly	−	shoe‡	——	(9a) car	+	car	car
(10) car	+	car	car	(10a) car	+	car	car
(11) car	+	car	car	(11a) rocky	+	rocky	rocky
(12) rocky	+	rocky	rocky	(12a) rocky	+	rocky	rocky
(13) rocky	+	rocky	rocky	(13a) rocky	+	——	——
(14) rocky	+	——	——	(14a) key	+	key	key
(15) keys	+	key	keys	(15a) dog	+	dog	——
(16) key	+	key	key	(16a) spoon	−	——	key
(17) key	+	key	key	(17a) Mommy	−†	——	——
(18) wawaw	−*	wawaw	wawaw	(1c) box	+	box	——
(19) fcahr	+	car	car	(2c) box	+	——	car
(20) fcahr	+	car	car				
(21) Mommy	−	——	shoe‡				

* Since no response could be elicited during Part 3 to the solid object, and since the parents do not agree as to the certainty of the name in previous handling, all identifications of this stimulus are doubtful.

† Much amusement.

‡ "Shoe" may be the judges' misinterpretation of *S*'s "thank you" in response to being given the pictures.

§ Misidentification due to sister's entry into the room.

This is *not* to say that *S* never had been exposed to pictures. There was a Japanese print on one wall of a room through which he frequently passed; a myriad of billboards fronted the highways on which he traveled frequently; a few times (six in all) he accidentally encountered a picture-book (which was gently withdrawn) or caught a glimpse of the label of a jar of baby food (these were normally removed or kept covered). (All these encounters were unaccompanied by instruction or naming-play.) Furthermore, one toy (a top) had pictures of elves on it and, accordingly, it was available for play only under strict supervision to prevent any naming in his presence; and a high chair had a decal of babies on it, which could be glimpsed (without parental comment) only when *S* was being placed in the seat.

The constant vigilance and improvisation required of the parents proved to be a considerable chore from the start—further research of this kind should not be undertaken lightly. By 19 mo. of age, the child began actively to seek pictures, and continuation of the constraints became both pediatrically and methodologically undesirable. Two incidents terminated this stage of the investigation: (a) *S* became aware of events on the TV set in the next room, managed to obtain a glimpse of the screen on which a horse was being depicted, and excitedly cried "dog"; (b) he squirmed around in his highchair about the same time, and, pointing to the decal, said "baby." It was evident that some form of parental response to such identifications would soon become unavoidable. The testing procedure was begun at this point.

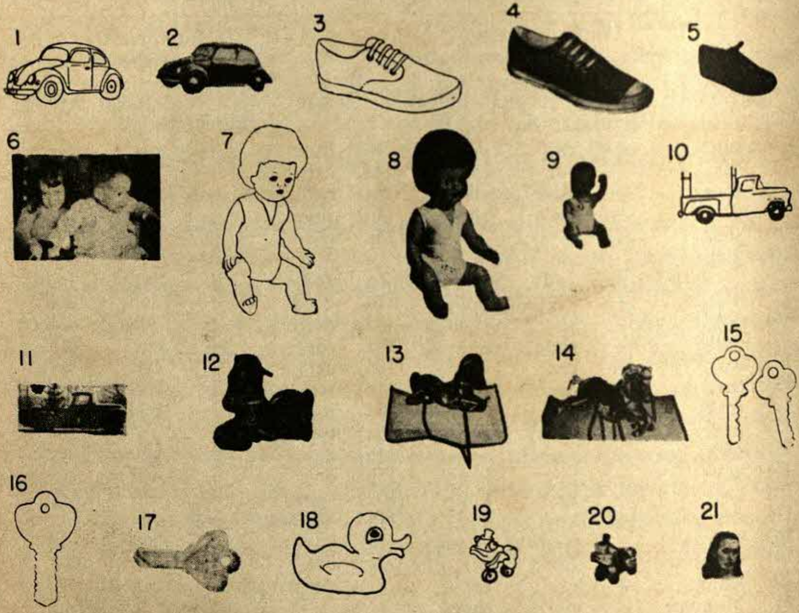


FIG. 1. PICTURES SHOWN *S* IN PART 1

Testing: Part 1. The set of 21 pictures listed in the first part of Table I, and shown in Fig. 1, was prepared on 3 × 5 in. cards. In all cases but one (No. 12), the series was so arranged that the outline-drawing of any object preceded any photograph of the same object; recognition could not, therefore, be made first from the photograph and then transferred to the drawing. The drawings were handed one at a time to *S*, a somewhat unsuccessful attempt being made to convert the test to an interesting game. Responses were obtained by tape-recording.

Part 2. Immediately after Part 1, *S* was given a large store of picture-books. For a period of one month, he had free (but monitored) access to still pictures, but motion pictures, TV, and picture-naming play still were completely avoided. (It was feared that motion pictures would provide a basis for attaching names or three-dimensional 'meanings' to the still pictures which do, after all, appear even in cinema-sequences.) A great variety of naming-reactions appeared during this period, but special pains were taken not to respond to any of these. Vocabulary-building by means of object-naming games continued during this month, but at special times, and with no pictures present. At the end of the month, the testing procedure of Part 1 was repeated with the new set of stimuli listed in the second part of Table I and shown in Fig. 2.[5]

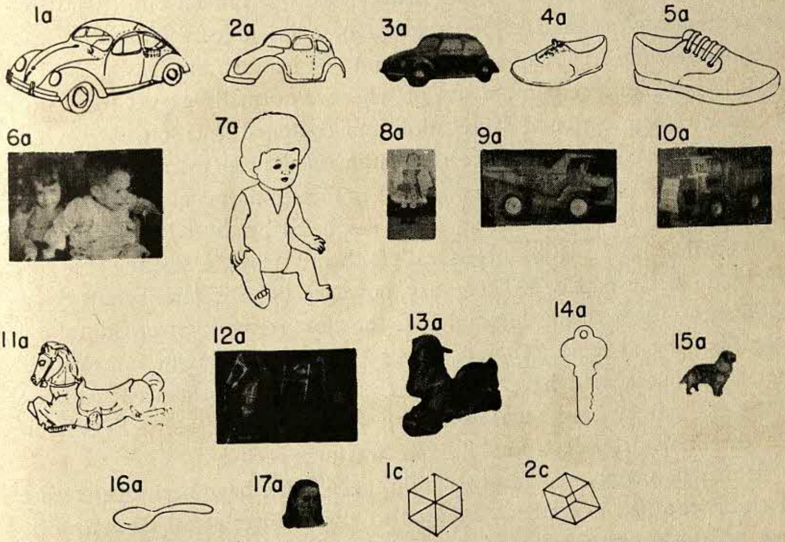


FIG. 2. PICTURES SHOWN *S* IN PART 2

Scoring. A set of objects, consisting of most of those whose pictures appeared in the two testing series, then was presented to *S*, and *his* naming-responses were tape-recorded. These provided two judges (*A* and *B* in Table I), who had not been present during the two testing sessions, with experience of the child's pronunciation. The judges then were told that those words would be used, in any order and with any number of repetitions, in the recordings of the two testing sessions, and their task was to determine the order of presentation of the objects, using only the child's responses as recorded during those sessions.[6] (This scoring procedure was undertaken separately by the two judges.)

---

[5] The object represented in 1*c* and 2*c* was a box constructed of rigid wire, approximately ⅛ in. in diameter, and 8 in. on a side. Introduced as a toy, it was involved in naming-games only during the month which elapsed between Part 1 and Part 2.

[6] The judges were permitted as many repetitions as desired. The first two presentations in Part 1 had been edited to remove extraneous chatter. Since the third

Results. Those pictures which were considered to be correctly identified by the parents (who did know the stimulus-series) are shown in Table I; the interpretations of each of the judges also are shown. If we consider only the line-drawings which both judges correctly identified in Part 1 (given eight possible names which, they were told, could appear in any order with any number of repetitions), the probability of a chance relationship is well under 0.01. The judges thus correctly identified objects from the *S*'s responses to the line-drawings.

It seems clear from the results that at least one human child is capable of recognizing pictorial representations of solid objects (including bare outline-drawings) without specific training or instruction. This ability necessarily includes a certain amount of what we normally expect to occur in the way of figure-ground segregation and contour-formation. At the very least, we must infer that there is an unlearned propensity to respond to certain formal features of lines-on-paper in the same ways as one has learned to respond to the same features when displayed by the edges of surfaces. "Ghost shapes," as Gibson has called them,[7] may be anemic, but they are by no means deceased. There may, however, be considerable ontogenetic difference in structuring or emphasis; the clear recognition of Stimulus 1c as a "box," and the uncertainty of response to Stimulus 2c, are certainly not what one would have expected from adult performance.[8] Although order-effects may have been responsible for the poor response to Stimulus 2c, the immediate identification of 1c was unexpected.

It may be maintained that this ability would not have been displayed by a child who had never been exposed to any pictures at all, and who had not had such extensive experience with patterned surfaces as had the present *S*. This may be true (although consideration of the inhomogeneities of the normal *apictorial* environment make it seem quite improbable to us), but, even if it is, the complete absence of instruction in the present case (the absence of 'association' between picture and represented object) points to *some* irreducible minimum of native ability for pictorial recognition. If it is true also that there are cultures in which this ability is absent, such deficiency will require special explanation; we cannot assert that it is simply a matter of having not yet learned the 'language of pictures.'

---

presentation proved to have elicited 'extraneous' responses which might be interpretable by judges as part of the series, editing of the remaining presentations was restricted to the elimination of overly-long interruptions; the residual chatter and gabble made judgment a difficult task.

[7] Gibson, *op. cit.*, 412.

[8] Hochberg and Brooks, *op. cit.*, 347.

QUALITATIVE CHANGES IN THE MEMORY-TRACES OF INCOMPLETE CIRCLES

By RONALD E. JOHNSON, Auburn University

The course of retention is characterized not only by the amount retained but also by the qualitative changes in the material. In 1922, Wulf reported that these qualitative changes progressed in the direction of good gestalten.[1] Although the experiments which followed Wulf's tended to support his findings, methodological errors were detected.[2] When these were eliminated, none of the investigations since 1941 has confirmed Wulf's results.[3]

More basic, however, than the question whether memory-changes progress in the direction of good gestalten is the question whether any qualitative changes occur in the memory-trace. A recent experiment, using the method of recognition for measuring forgetting, yielded results which indicate that no qualitative changes occur in the memory-trace.[4]

The major proving ground for the investigations of these changes has been the recognition of visual figures; and the most popular stimulus-figure has been the incomplete or broken circle. In at least nine investigations, this figure has been used as the stimulus-object.[5] The results of these investigations are inconsistent. In some of them, closure occurred; in others, the gap was emphasized; and in still others, different combinations of closure and emphasis of the gap were reported. The re-

---

* Received for publication April 20, 1961.

[1] Friedrich Wulf, Ueber die Veränderung von Vorstellungen (Gedächtnis und Gestalt), *Psychol. Forsch.*, 1, 1922, 333-373.

[2] Cf. R. E. Johnson, A quantification and measurement of three qualitative changes in the recall of complex verbal materials, Unpubl. doctoral dissertation, Ohio State Univ., 1960.

[3] Erich Goldmeier, Progressive changes in memory traces, this JOURNAL, 54, 1941, 490-503.

[4] Irvin Rock and Phima Engelstein, A study of memory for visual form, this JOURNAL, 72, 1959, 221-229.

[5] F. W. Irwin and M. A. Seidenfeld, The application of the method of comparison to the problem of memory change, *J. exp. Psychol.*, 21, 1937, 363-381; D. O. Hebb and E. N. Foord, Errors of visual recognition and the nature of the trace, *ibid.*, 35, 1945, 335-348; F. H. George, Errors of visual recognition, *ibid.*, 43, 1952, 202-206; N. G. Hanawalt, The method of comparison applied to the problem of memory change, *ibid.*, 43, 1952, 37-42; J. C. Crumbaugh, Temporal changes in the memory of visually perceived form, this JOURNAL, 67, 1954, 647-658; J. B. Carlson and C. P. Duncan, A study of autonomous change in the memory trace by the method of recognition, this JOURNAL, 68, 1955, 280-284; E. L. Walker and Joseph Veroff, Changes in the memory-trace for perceived forms with successive reproductions, this JOURNAL, 69, 1956, 395-402; Lawrence Karlin and Glen Brennan, Memory for visual figures by the method of identical stimuli, this JOURNAL, 70, 1957, 248-252; H. James, Guessing, expectancy and autonomous change, *Quart. J. exp. Psychol.*, 10, 1958, 107-110.

sults of these experiments are consistent, however, when gaps of similar sizes were used, *i.e.* the same qualitative changes were reported—and this occurred in spite of the fact that widely different procedures were used. The trends exhibited are as follows: (1) when circles with gaps of 2° to 18° are used, the gaps are emphasized; (2) when gaps of 20° to 40° are used, slight emphasis and then closure results; and (3) when circles with gaps of more than 50° are used, closure follows. James interprets the inconsistency of his results to mean that "the cues determining judgment differ from one size gap to the other."[6] This is in close agreement with Walker and Veroff's earlier statement that "it would require an empirical examination of a wide variety of circles with various gaps, to determine when to expect sharpening and when to expect closure."[7]

The object of the present study is to test the hypothesis that the qualitative changes which occur with incomplete circles are functions of the size of the gaps. Specifically, does closure occur with large gaps and emphasis or sharpening with small?

Materials. Broken circles with gaps of 10°, 15°, 20°, 30°, 40°, 50°, 60°, and 70° were used for stimulus-figures. All the gaps were centered at the 12 o'clock position. The circles had diameters of 4.6 cm. and were drawn in India ink on 15-cm. squares of white posterboard. Each inked line was 1-mm. wide.

The figures used for recognition were drawn for each 2½°-interval from 2½° to 90°. All recognitive series of cards for each stimulus-card of 30° and above contained 17 cards ranging from 20° below the stimulus-figure to 20° above it. The recognitive series for the 10°, 15°, and 20° gaps all began at 2½° and continued through 17½°, 27½°, and 37½° respectively. The number of recognitive cards in these three series thus ranged from 7 to 15. Each series was put in consecutive order.

Subjects. A total of 675 undergraduate students was used in groups of 25 to 36 each. To eliminate the memory-distortion of repeated testing, individual *S*s made recognitive judgments on only one occasion.[8] Recognitive judgments of different groups were made after 5 min., 1 week, and 2 weeks. There were thus a total of 24 different groups of *S*s.

Procedure. Each *S* was shown the figure for 5 sec. and was told to "remember the figure as exactly as you can." The sitmulus-card was then handed to the *S*, who was allowed to hold the card at any preferred distance from himself. The *S*s were not told there was to be a later recall-period.

When the *S*s returned to the experimental room for their recall, they were told individually to "select the figure which is identical to or most nearly like the figure you have already seen." As the instructions were given to the *S*s, the recognitive series of cards was placed before them. The series was opened at different cards for different *S*s. For each *S*, both the card at which the series was opened and the card chosen were recorded.

The total number of choices which had gaps smaller than the original stimulus-card was equal to the possible choices which had gaps larger than the original card.

---

[6] James, *op. cit.*, 110.

[7] Walker and Veroff, *op. cit.*, 402.

[8] Hanawalt, Memory trace for figures in recall and recognition, *Arch. Psychol.*, 31, 1937 (No. 216), 1-89.

Since the series was opened at different cards for different *Ss*, the number of *Ss* who had gaps smaller than the stimulus-figure by a specific number of degrees was counterbalanced by the same number of *Ss* who had gaps larger than the stimulus-figure by the same specific number of degrees. Controlling the position of the correct card in the total series and the degree of gap in the opening card was necessary to correct for the tendency of *Ss* to select a card which errs in the direction of the error of the opening card of the recognitive series.[9]

Results and discussion. Table I presents the mean choices for the different stimulus-figures at each of the time-intervals. Bartlett's test demonstrated beyond the 1% level that the variances were non-homogeneous. To test the possibility that the non-homogeneity was due to the restricted recognitive choices of the 10°, 15°, and 20° intervals, Bartlett's test was reapplied to the recognitive intervals from 30° to 70°. The lack of statistical significance for the second test of homogeneity made it possible to assume that the variances of these five groups were homogeneous. A two-way analysis of variance for the groups from 30° to 70° was then computed. To facilitate the interpretation, 72 *Ss* were randomly removed from the groups to make them equal in number. A total of 375 *Ss* was then left for the analysis.

TABLE I

MEAN CHOICE AT DIFFERENT TIME-INTERVALS AS A FUNCTION OF GAP-SIZE ($N=675$)

Recognitive time-intervals

Size of gap	5 min.			1 week			2 weeks		
	M	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*
10°	10.18	2.60	28	9.48	4.94	29	9.33	4.36	26
15°	17.69	4.01	27	18.70	6.94	27	18.48	7.29	28
20°	22.21	7.45	34	24.54	8.44	27	22.23	8.34	28
30°	34.35	9.84	27	34.10	8.84	25	34.57	9.82	29
40°	40.83	8.10	27	40.28	11.40	27	39.60	13.08	31
50°	54.63	8.40	27	51.06	8.91	26	54.31	8.50	36
60°	58.52	9.76	27	54.90	12.85	25	58.75	11.90	31
70°	69.02	10.80	28	59.44	8.84	27	65.89	12.38	28

If different kinds of qualitative change are a function of the size of the gaps, the interaction-variance would be expected to be statistically significant. The lack of significance indicates that different kinds of qualitative change did not occur for the various gap-sizes included in the analysis.

The significant *F*-ratio for the three time-intervals suggests that the same kind of qualitative change may have occurred for all gap-sizes ($F =$

---

[9] Carlson and Duncan, *op. cit.*, 280-284.

3.81, $df = 2$, $p < 0.05$). The possibility of like qualitative changes was explored further by making one-way analyses of variance for the gap-sizes between 30° and 70°. Individual tests of homogeneity indicated that the variances for each of the gap-sizes can be assumed to be homogeneous. The only one-way analysis of variance which was statistically significant beyond the 5% level was the 70° interval. The F-ratio of 5.46 was significant at the 1% level. The differences among the time-intervals for the different gap-sizes were assessed also by Kruskal and Wallis' ranked one-way analysis of variance. As before, the only statistically significant difference was for the 70° interval ($H = 9.2$, $df = 2$, $p < 0.02$).

The non-significance of seven out of the eight gap-intervals tested and the smallness of the F-ratio for the five gap-intervals tested in the two-way analysis of variance thus give only minimal support to the hypothesis that the same kind of qualitative change occurred for all gap sizes. More support for this interpretation is evidenced by the mean changes of the 10°, 15°, and 20° groups. These last three groups, which were not included in the two-way analysis, had mean changes which ran counter to the changes exhibited by the 30° to 70° groups.

To check Carlson and Duncan's finding that *S*s will select a card which errs in the direction of the opening card of the recognitive series, the size of the opening card was related to the mean recognitive card.[10] The correct card was chosen by 35 of the 421 *S*s in the 30° to 70° groups. Of the 194 presented with minus opening cards, 93 chose minus cards. Of the 192 presented with positive opening cards, 120 chose positive cards. The Chi-square of 4.30 with 1 *df* was significant at the 5% level. The mean choice of those presented with negative cards was −0.325; for those presented with positive cards, the mean choice was +0.592. A t-ratio of 2.07 was also significant at the 5% level. The difference in recognitive choices of those who were presented negative top cards and of those who were presented top cards was thus nearly 2½°. Carlson and Duncan's finding was confirmed.

Summary. The results of this study give no support to the hypotheses that closure occurs with large gaps and sharpening occurs with small gaps. They furthermore give no support to the gestalt hypotheses that perceptual and memorial closure will occur for the incomplete circle.

---

[10] The card series was numbered from −8 to +8; the minus cards are those with smaller openings than the standard, while the plus cards are those with larger openings. Each integral difference represents a difference of 2½°.

Only two broken-circle stimulus-figures (10° and 40°) showed continuous memory change in the same direction. Statistical significance was shown in neither case. Memory change for the 70° gap was the only one which reached statistical significance. If the data for the 70° gap are excluded, the remaining data strongly suggest that qualitative changes in figural memories do not occur.

PROBABILITY-MATCHING IN THE PIGEON

By DONALD H. BULLOCK and M. E. BITTERMAN, Bryn Mawr College

Probability-matching is a well-known phenomenon in man, but the only infrahuman animal in which it has been found thus far is the fish *Tilapia macrocephala* (commonly known as the African mouthbreeder). Matching in the mouthbreeder first appeared in some preliminary work with 70:30 problems (both visual and spatial) under conditions which insured that one of the alternatives would be reinforced on exactly 70% of trials and the other on exactly 30% of trials,[1] and further evidence of matching was found in subsequent work with 80:20, 70:30, 60:40, and 50:50 problems (in which visual and spatial cues were confounded).[2] Studied under analogous conditions, rats and monkeys maximize; that is, they tend to choose the more frequently reinforced alternative on about 100% of trials.[3]

In the experiment to be reported here, evidence of probability-matching was sought in the pigeon, an animal intermediate in complexity between fish and mammal. The experiment is one of a series on the learning of pigeons under conditions in which comparative studies of higher and lower forms have yielded contrasting outcomes.

Subjects. The *Ss* were 10 experimentally naïve adult male White Carneaux pigeons. They were maintained at 80% of their free-feeding body-weights by the technique of Bullock, Roberts, and Bitterman.[4]

Apparatus. The *Ss* were tested in a small chamber which was set into a sound-

---

* Received for publication May 9, 1961. This research was supported by the Office of Naval Research under terms of Contract Nonr 2829(01). The assistance of William A. Roberts and Rachel Tucciarone is acknowledged with thanks.

[1] M. E. Bitterman, Jerome Wodinsky, and D. K. Candland, Some comparative psychology, this JOURNAL, 71, 1958, 94-110.

[2] E. R. Behrend and M. E. Bitterman, Probability-matching in the fish, this JOURNAL, 74, 1961, 542-551.

[3] Bitterman, Wodinsky, and Candland, *op. cit.*, 103-108; Allen Parducci and James Polt, Correction *vs.* noncorrection with changing reinforcement schedules, *J. comp. physiol. Psychol.*, 51, 1958, 492-495; W. A. Wilson, Jr., and A. R. Rollin, Two-choice behavior of rhesus monkeys in a noncontingent situation, *J. exp. Psychol.*, 58, 1959, 174-180; W. A. Wilson, Two-choice behavior of monkeys, *ibid.*, 59, 1960, 207-208; D. R. Meyer, The effects of differential probabilities of reinforcement on discrimination learning by monkeys, *J. comp. physiol. Psychol.*, 53, 1960, 173-175. Some results reported by W. K. Estes (Of models and men, *Amer. Psychologist*, 12, 1957, 609-616) have been taken as evidence of matching in the rat, but the data for individual *Ss* do not support such a conclusion. For details, see Bitterman, Wodinsky, and Candland, *op. cit.*, 107 (ftn. 18).

[4] D. H. Bullock, W. A. Roberts, and M. E. Bitterman, Techniques of housing and maintaining a large pigeon colony, *J. exp. Anal. Behav.*, 4, 1961, 285-286.

attenuating picnic chest. The front wall of the chamber contained two Lehigh Valley Electronics pigeon-keys separated by the aperture of the Gerbrands pigeon-magazine. Experimental contingencies were programmed automatically, and responses were recorded on a Gerbrands four-channel polygraph.

Procedure. After receiving at least 1 hr. of *VI-1 min.* and 1 hr. of *VI-2 min.* magazine-training, each bird was conditioned to peck a key by a combination of shaping and baited-key techniques. This pretraining was conducted in the two-key apparatus with the right-hand key lighted red and the left-hand key unlighted. Once pecking was established, a controlled operant or discrete-trials procedure was introduced in which each reinforcement (4-sec. access to grain) was followed by a 6-sec. timeout (*TO*) or intertrial interval. During the intertrial interval, the key-light was turned off and the chamber was dark, because the only other source of illumination was the magazine-light, which was on only during reinforcement. Both keys were used for this discrete-trials training; they were presented individually in balanced order, each on half the trials.

The data to be reported here were obtained after some preliminary work with a 70:30 spatial problem, in which both keys were red and the right-hand key was reinforced on 70% of trials. The performance of the birds under those conditions was not sufficiently consistent to warrant definite conclusions: primarily, it seemed, because of the persistence of rather strong initial preferences for one or the other of the keys. In the hope of circumventing that difficulty, a visual problem was substituted for the spatial one. On each trial, one key was red and the other yellow, the spatial arrangement of the two colors being varied systematically from trial to trial in accordance with Gellermann's rules.[5] Throughout the subsequent training of the birds, position was nondifferentially reinforced, but (except for a series of trials on a 50:50 problem in which there was no differential reinforcement whatever) the probability of reinforcement for response to one color was greater than the probability of reinforcement for response to the other, the two probabilities summing always to unity.

In the first phase of the experiment, 5 birds (*PM*s 1, 3, 7, 9, and 10) were given 32 sessions of 100:0 reinforcement with the red key as the 100% key. Five other birds (*PM*s 2, 4, 5, 6, and 8) were given 8 sessions of 100:0 reinforcement with the red key as the 100% key, followed by 23 sessions of 70:30 reinforcement with the red key as the 70% key. The distribution of reinforcements on the yellow key was random, with the restriction that there must be at least 1 in each block of 5 trials. All 10 birds then received 17 sessions of 0:100 reinforcement, with response to the yellow key now being reinforced 100% of the time.

In the second phase of the experiment, the *S*s were regrouped; *PM*s 4, 6, 7, 8, and 9 were given 23 sessions of 20:80 reinforcement, with the yellow key (formerly 100%) as the 80% key; *PM*s 1, 2, 3, 5, and 10 were given 23 sessions of 40:60 reinforcement, with the yellow key as the 60% key. All 10 birds then were given 30 sessions of 50:50 reinforcement, with the exception of *PM* 3, which died after the sixteenth 50:50 session, and of *PM*2, which died after the eleventh 50:50 session. In the 20:80 problem, there was 1 red reinforcement in each block of 5 trials, while in the 40:60 problem there were 2. In the 50:50 problem, Gellermann's orders were employed to determine which color would be reinforced on each trial.

---

[5] L. W. Gellermann, Chance orders of alternating stimuli in visual discrimination experiments, *J. genet. Psychol.*, 42, 1933, 206-208.

In each phase of the experiment, there were 40 trials per day, and a 'guidance' procedure was used. If, on any trial, the bird pecked the reinforced key, both key-lights went out, and it received access to grain for 4 sec., after which there was a 6-sec. *TO*. If the bird pecked the unreinforced key, both key-lights went out, but there was no reinforcement, only a 6-sec. *TO;* then the correct key alone was lighted and the bird reinforced for response to it, after which there was another *TO* of 6 sec. Guidance was used because prior experimentation with it had yielded differences between rat and fish, the former 'maximizing' and the latter 'matching,' while nonguidance had yielded maximizing in both species.

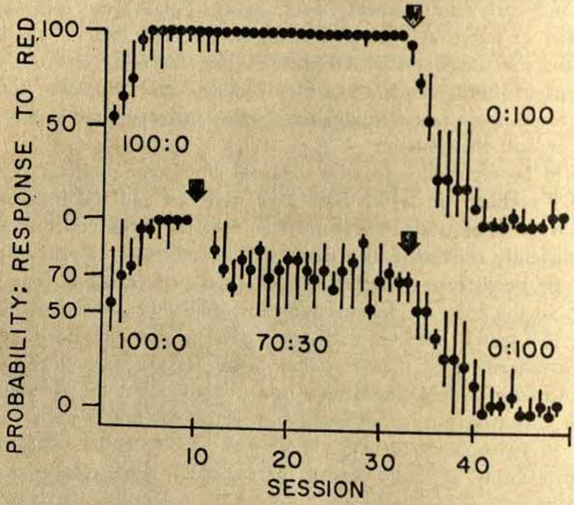


FIG. 1. THE PERFORMANCE OF THE ANIMALS IN THE FIRST PHASE OF THE EXPERIMENT
(The arrows show the points at which conditions were changed.)

Results. The measure of performance to be considered is the percentage of the trials in any session on which *S* pecked the red key first. For purposes of consistency, this measure is used throughout, whatever the probability of reinforcement for response to the red key.

Fig. 1 shows the median score of each group in each session during the first phase of the experiment. The range of scores in each session is indicated by a vertical line. The upper panel presents the data for the group which had 100:0 reinforcement (red key = 100%) for the first 32 sessions, and 0:100 (red key = 0%) for the next 17 sessions. The lower panel presents the data for the group which had 100:0 reinforcement (red key = 100%) for the first 9 sessions, 70:30 reinforcement (red key = 70%) for the next 23 sessions, and 0:100 reinforcement (red key = 0%) for the last 17 sessions. The data show clearly that both groups approximated an asymptotic performance of 100% on the red key after nine sessions of 100:0 reinforcement. The group that was continued on 100:0 maintained this performance, while the group that was shifted to 70:30 produced an asymptotic performance approximating 70% choice of red key. When the 0:100 condition was introduced, both groups shifted (rapidly and at about the same rate) to a near-zero level of response to the red key.

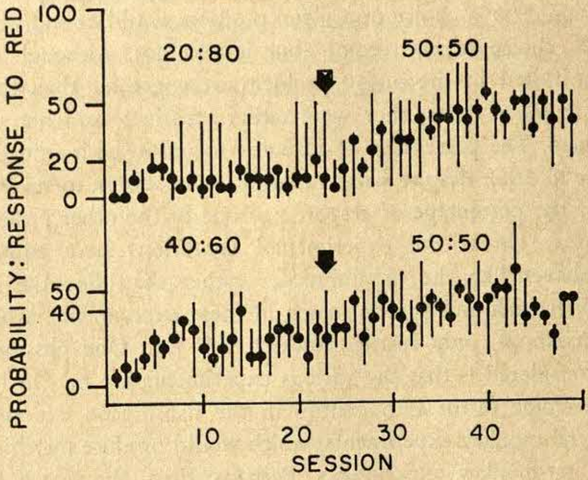


FIG. 2. THE PERFORMANCE OF THE ANIMALS IN THE SECOND PHASE OF THE EXPERIMENT
(The arrows show the points at which conditions were changed.)

Fig. 2 shows the median scores of the two groups for each session in the second phase of the experiment, with the range of scores in each session again being indicated by a vertical line. The upper panel presents the data for the group which was given 20:80 reinforcement (red key = 20%) followed by 50:50 reinforcement. The lower panel presents the data for the group which was given 40:60 reinforcement (red key = 40%) followed by 50:50 reinforcement. The curves give clear evidence of increase in the proportion of red choices as the probability of reinforcement for response to red was increased from 0% (in the previous phase of the experiment) to 20% or to 40%. The increase was greater in the 40:60 problem than in the 20:80 problem, although there was some overlapping between the two groups, and in neither case did the probability of response to red quite reach the probability of reinforcement in the time allotted (the level of response to red was only about 15% during the last 10 sessions on the 20:80 problem, and only about 30% during the last 10 sessions on the 40:60 problem).

When the ratio was changed to 50:50, the probability of response to red rose in both groups to about the 50% level. Such performance might be an artifact of the development of spatial preferences—with each color reinforced on 50% of trials, and each color appearing equally often in each position, 100% choice of a given position would be indistinguishable from 50% choice of each color—but it was not. Although the spatial biases manifested in the 50:50 problem were greater than those which appeared at other ratios, they were rather transient, and not all animals showed them. The percentage of response to red in each animal remained quite close to 50% despite wide variation from session to session in some animals of the percentage of response to one or the other position.

Discussion. Under the experimental conditions here employed, the pigeon behaves like the mouthbreeder rather than like the rat or the monkey. Why fish and pigeon match in these experiments, while rat and monkey maximize, only further research will tell. One possibility which must be considered is that the various experiments are not, in fact, analogous—that some factor is operating in the fish-pigeon experiments, but not in the rat-monkey experiments, which would produce matching in truly analogous rat-monkey experiments. Perhaps, then, the search for matching in rat and monkey should be taken up again with renewed vigor. Another possibility to be considered is that animals differ in the way in which they learn or the way in which they utilize the products of learning. Systematic comparison of the species in question under a variety of experimental conditions might reveal a meaningful pattern of similarities and differences in terms of which any particular difference could be understood. It may be of interest to note in this connection that, in conventional experiments on extinction after partial reinforcement and on progressive improvement in habit-reversal, the behavior of the pigeon is more like that of the rat than like that of the fish.[6]

Summary. Pigeons were trained in a two-key apparatus with access to food as reward for correct choice. The keys differed in color (red vs. yel-

---

[6] Roberts, Bullock, and Bitterman, Resistance to extinction in the pigeon after partial reinforced instrumental training under discrete-trials conditions, this JOURNAL, in press; Bullock and Bitterman, Habit reversal in the pigeon, *J. comp. physiol. Psychol.*, in press.

low), and the spatial arrangement of the two colors was varied in quasi-random fashion from trial to trial, each color appearing equally often in each position. Trials were discrete (each correct response was reinforced, and each reinforcement was followed by a time-out), and a guidance procedure was used which made it possible to fix exactly the number of reinforcements for response to each color in each position (after each incorrect response, the correct key alone was lighted and the animal reinforced for response to it). Under these conditions, probability-matching appeared in the pigeon; that is, the proportion of responses to a given color varied directly with the probability of reinforcement. This phenomenon has been demonstrated in the fish, but it has failed to occur in what appear to be analogous experiments with rat and monkey.

DECISION-MAKING IN AN ACTUAL GAMBLING SITUATION

By ROBERT F. MUNSON, V.A. Hospital, Louisville, Kentucky

In studies of decision-making under conditions of uncertainty, the expected-value model, or some variation of it, has been used to predict choice.[1] It has been asserted, however, that the variance of an offer as well as its expected value (*EV*) may be important in influencing decisions,[2] and, furthermore, that income-level is an important factor in choice of offers.[3] McGlothlin's findings support the thesis that variance influences the choice of bets.[4] McGlothlin, who used an actual gambling situation (race-horse betting), did not, unfortunately, analyze his data according to the choices made by groups of different income-level.

In laboratory studies of decision-making, gambling games usually are used. The *S*s know that they are participating in an experiment, and they are given a stake to play with or are paid for participating. Seldom are the *S*s in danger of losing any of their own money. It is possible, however, that people may use different strategies and make different choices when they are using their own money and are unaware that their decisions are being scrutinized. The present study represents an effort to answer four questions about decision-making in an actual gambling situation in which the participants used their own money and were unaware that their decisions were being examined. These questions are: (1) Will the choices made by the various groups be consistent enough to indicate stability of decision? (2) Will the *EV* of an offer be related to choices of it? (3) Will

---

[*] Received for publication February 1, 1961. This paper is adapted from a Doctoral dissertation submitted to the University of Kentucky. The writer is indebted to Dr. Richard Griffith for help and advice.

[1] W. D. Edwards, Probability-preferences among bets with different expected values, this JOURNAL, 67, 1954, 56-67; The prediction of decisions among bets, *J. exp. Psychol.*, 50, 1955, 201-214; Frederick Mosteller and Philip Nogee, An experimental measurement of utility, *J. polit. Econ.*, 59, 1951, 371-404; Alvin Sodel, Philburn Ratoosh, and J. S. Minas, Some personality correlates on decision making under conditions of risk, *Behav. Sci.*, 4, 1959, 19-26.

[2] Maurice Allais, Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'ecole Americaine, *Econometrica*, 21, 1953, 503-546; Gerhard Tintner, A contribution to the non-static theory of choice, *Quart. J. Econ.*, 56, 1942, 274-276.

[3] Milton Friedman and L. J. Savage, The utility analysis of choices involving risk, *J. polit. Econ.*, 56, 1948, 279-304; Ratoosh and Minas, *op. cit.*, 19-26.

[4] W. H. McGlothlin, Stability of choices among uncertain alternatives, this JOURNAL, 69, 1956, 604-615.

the risk of an offer (as measured by its variance) influence choice? (4) Will groups with different levels of income show a preference for different offers according to their risks?

Subjects. The *S*s were participants in a gambling game conducted at various public carnivals in a mid-western city with a population of over 700,000.

Apparatus and procedure. In a booth located on the midway of the carnival, nine silver dollars, painted a bright red color on one side, were placed in a plastic container and after being thoroughly shaken were tossed, in plain view of the participants, on a small table. The red heads appearing were rapidly counted and the number announced to the participants in a loud voice. This procedure was repeated over and over during the entire course of the evening at a pace limited only by the speed of the croupier.

On a table immediately in front of the bettors was a board divided into 20 slots. In the 20 slots were inserted 2 sets of the numbers 0-9 on which the bettor placed his nickel. The board was so constructed that the numbers were movable in order that their positions could be randomized. The positions of the numbers on the board were changed at 30-min. intervals according to a predetermined schedule, since, in a pilot-study, the spatial location of a number seemed to be a significant factor in how frequently it was played.

TABLE I

THE TWO SCHEDULES OF PRIZES

	Equal expected values			Different expected values		
No. of reds	Prize	*EV**	Variance	Prize	*EV**	Variance
0	\$12.75	0.0250	31.883	\$10.20	0.0300	20.445
1	1.35	0.0254	3.385	2.20	0.0104	8.743
2	0.30	0.0248	0.801	0.25	0.0288	0.588
3	0.10	0.0254	0.309	0.05	0.0336	0.137
4	0.05	0.0254	0.199	0.05	0.0254	0.199
5	0.05	0.0254	0.199	0.10	0.0131	0.448
6	0.10	0.0254	0.309	0.10	0.0254	0.309
7	0.30	0.0248	0.801	0.40	0.0183	1.324
8	1.35	0.0254	3.385	0.50	0.0403	0.522
9	12.75	0.0250	31.883	2.20	0.0460	0.987

* All expected values are negative.

The winners were paid according to the schedule shown in Table I with equal *EV*s used one night and different *EV*s the next. A flip of a coin decided which set of prizes was used first at the first location and also determined the set of prizes to be used on all other nights in order to have a counterbalanced design.

The board with the numbers inserted was mounted on hinges and attached to a trough divided into 20 slots corresponding to the numbers. These slots were so placed that all nickels played on any one number fell into its corresponding slot, where they were retained until collected and counted at the end of each 30-min. interval for a period of 4 hr. This method of conducting the game and collecting the data interfered minimally with the actual playing and prevented the bettors from knowing that they were actually participating in an experiment.

The game was conducted at five different locations. Two of these were in neighborhoods of higher median yearly income (*HIL*) and three were in neighborhoods of lower yearly income (*LIL*). Both of the *HIL*s and one of the *LIL*s had carnivals on two consecutive nights and were located five blocks apart. The data colleceted at these two *LIL*s were treated as if they had been obtained at one *LIL*.

Results. To test for consistency between the two different locations within the same level of income when the same schedule of prizes was used, the offers were ranked at each location by their popularity during each of the eight 30-min. periods. The coefficient of concordance (*W*) was found for the ranks, and the *W*s were transformed, to Chi-squares all of which were significant beyond the 1% level. This result indicates that the groups were highly consistent in their pattern of choice when the game was conducted under the same conditions. The first question asked is answered in the affirmative.

The second question asked is whether choice of offer is related to *EV*. In an effort to answer this question, rank-order of the offers arranged by popularity were compared with the expected rank-order of the offers by size of *EV*. At both the *HIL*s and the *LIL*s, and when all the locations were combined, the rank-order correlations between these sets of ranks were significant beyond the 5% level. Apparently, the *EV* is an important factor in choice of offers.

The third and fourth questions asked are whether the risk of any offer is related, in any way, to choice, and whether difference in income is related to a preference for offers according to the degree of risk involved. To answer these questions, the total number of nickels played on the four offers having the largest variance (0, 9, 1, 8) were compared with the total number played on the four offers having the lowest variance (3, 4, 5, 6) under the condition of equal *EV*. A plus was recorded when the total played on the low-variance offers was greater than the total played on the high-variance offers, and a *minus* when smaller. Thus, a *plus* or *minus* sign was obtained for each 30-min. period during which the game was conducted (a total of 32 signs). At the *HIL*s, 11 *pluses* and 5 *minuses,* and at the *LIL*s, 11 *pluses* and 5 *minuses,* were obtained. That is, there was a tendency of the same strength in both groups to favor the low-variance offer.

Discussion. The finding that the *EV* of an offer was related to its popularity is contrary to most of the previous results in this area and suggests that there is an essential difference between the strategy used by people in a laboratory setting and in a hot, humid, exciting, and highly disordered carnival setting. It suggests also that people can act rationally in terms of the *EV*-model when a rational course is open to them and it is important for them to do so. The *S*s in this study were using their own money and were trying, it is safe to assume, to win some money.

The results reported by McGlothlin for a free gambling situation support the thesis that the risk of an offer is an important factor in influencing decision-making under conditions of uncertainty. Since the groups in the present study showed an over-all consistency in their choices in terms of rank-order popularity of offers, even when the *EV*s of the offers all were the same, it is possible that the variance of an offer may have had some influence and acted as a stabilizer under these conditions. That the risk of an offer did influence some participants was evidenced by comments frequently overheard by the workers in the booth such as "Let's go for broke" and "Let's play it safe." The lack of over-all consistency in terms of preference for certain variances may perhaps be traced to the influence of the large payoffs at peak playing hours. For example, at one of the *LIL*s, at least one of the participants won $12.75 in three of the last four periods. This outcome resulted in a tremendous play on the high-variance offers during those periods, which shows up in the reversal of signs for these last four periods. Thus, while it is impossible to say on the basis of a statistical analysis that a preference for certain variances exists, there was a decided tendency to favor the low-variance offer.

Summary. A genuine gambling game was conducted at a series of public carnivals held at neighborhoods with two different levels of income. The *EV*s of the offers were adjusted. On one night, all offers had the same *EV*s, while, on a second night, all the offers had different *EV*s. The results indicate that: (1) the groups were highly consistent in their choice of offers; (2) the *EV* of an offer was significantly related to the groups' choice of it; (3) there was a tendency by all the groups to avoid the high-risk offers and to prefer the small but frequent rewards of the low-risk offers; and (4) the high and low income groups did not differ in their preference for offers according to the degree of risk involved.

A VERBAL MEASURE OF THE PERSPECTIVE ATTITUDE

By V. R. Carlson and E. P. Tassone,
National Institute of Mental Health

In several previous experiments it has proved useful to interpret judgments of size-constancy, that have been obtained experimentally, in terms of the perspective attitude—defined as the implicit assumption on the part of *O* that the visual appearance of the magnitude of an object diminishes with increasing distance of the object.[1] It was postulated that this attitude represents a convention for communicating subjective experience rather than a functional relationship between perceived size and object-size in the natural three-dimensional environment. If this argument is correct, it should be possible to demonstrate the perspective attitude in a purely verbal way, without reference to actual objects and distances at all, and that is the primary purpose of the present study. The argument also would require a correlation between an independent measure of the perspective attitude and size-constancy performance, depending upon the specific instructions used regarding size. To test this expectation, judgments of size-constancy were obtained from the same *O*s.

Method: Observers. The *O*s were 30 college students (13 men and 17 women) ranging in age from 18–26 yr. None had prior experience either with the verbal or the size-constancy task.

Measure of verbal size-distance association. Twelve items of the form used in semantic-differential scales were constructed.[2] In this case, however, all the words were adjectives. Each item was typed separately on a slip of paper and consisted of an adjective on the right, one on the left, seven underlined spaces between the two, and a third centered above the middle space.

The items making up the test were as follows, according to one of the orders used: (1) long—rectangular—short; (2) nearby—bright—distant; (3) large—far away—small; (4) light—small—heavy; (5) dim—distant—bright; (6) nearby—large—far away; (7) sharp—circular—blunt; (8) small—nearby—large; (9) nearby—dim—distant; (10) heavy—large—light; (11) far away—small—nearby;

---

* Received for publication April 7, 1961.

[1] V. R. Carlson, Overestimation in size-constantcy judgments, this Journal, 73, 1960, 199-213; Effects of sleep-deprivation and chlorpromazine on size-constancy judgments, this Journal, 74, 1961, 552-560; Size-constancy judgments and perceptual compromise, *J. exp. Psychol.*, 63, 1962, 68-73.

[2] C. E. Osgood, G. J. Suci, and P. H. Tannenbaum, *The Measurement of Meaning*, 1957, 81.

and (12) dim—nearby—bright. The words written first and third in each case were the end-words.

Each *O* was tested individually, and the items were presented one at a time. *O*'s task was to decide whether the center-word possessed a connotative association with the word on the right or the word on the left and to indicate the degree of association. The mechanics of the task were initially explained using a dummy item with the words replaced by the letters *A, B, C.* It was emphasized to *O* that the association was expected to be intuitive or impressionistic rather than logical, rational, or concrete. *O* then was presented with the first item, "long—rectangular—short," and told:

> Now you see, logically, one could argue that 'rectangular' does not mean 'long' anymore than it does 'short.' One could equally well argue that it means both or neither; but you may have the impression, without any particular reason for it, that one of those two words seems more generally associated with 'rectangular' than the other. Choose that one, and if it seems to be a very strong association check the space nearest the word. If the association seems not so strong, check one of the spaces closer to the middle, and if you get absolutely no impression of any association either way check the space in the middle.[3]

The order of the remaining items varied from one *O* to another. The right-left location of the words within items varied both among *O*s and among items to control for any right-left preference. The items were distributed to prevent, as much as possible, those with the same words occurring in succession.

The four items of size-distance content (Items 3, 6, 8, and 11) were the essential ones for the present purposes. Each was scored on a scale from 1–7 in the direction consistent with association of 'far away' with 'small,' and 'nearby' with 'large.' The arithmetic mean of these four values constituted *O*'s verbal size-distance associative score, a higher score referring to a higher degree of association in the far away-small, nearby-large direction.

The rest of the items were primarily fillers, but as a matter of interest four (Items 2, 5, 9, and 12) were constructed to measure a possible brightness-distance association. They were scored similarly to the size-distance items, in the direction of associating 'distant' with 'dim,' and 'nearby' with 'bright.'

All the *O*s completed the verbal scale before making any judgments of size-constancy.

Size-constancy task. Essentially, *O* was required to adjust a variable triangle at a distance of 10 ft. to match various sizes of a standard triangle located at a distance of 40 ft. *O* was allowed free binocular regard under conditions of nonreduction: Instructions regarding apparent size were given on the first day; instructions regarding perspective size on a second day; and instruction regarding objective size on a third day. The procedures, instructions, and stimulus-conditions were the same as those described more extensively in the previous reports.

Results. The verbal-association scores were markedly concentrated toward the high end of the 7-point scale. The mean-value for size-distance association was 5.9, with only one of the 30 *O*s producing a score as low as the neutral point, 4.0. The mean for brightness-distance association was 5.6, with three scores falling at or below 4.0. 'Small' and 'dim' were quite unequivocally associated with 'far away' or 'distant' in preference to 'nearby'; 'large' and 'bright,' with 'nearby' in preference to 'far away' or 'distant.'

---

[3] Most *O*s selected 'long' and checked either the space nearest to it or the one next removed. It appears in retrospect that these instructions may have encouraged the frequent use of the extreme category. We discern no bias of the instructions, however, with respect to the direction of *O*'s choice between the two alternatives.

The mean variable-to-standard size-ratios were 0.96, 1.31, and 1.51, respectively, with instructions for the apparent size, objective size, and perspective size. These values agree closely with the averages of previous groups, who have regularly shown size-constancy with instructions for apparent size, overconstancy with instructions for objective size, and marked overconstancy with instructions for perspective size.

TABLE I

CORRELATIONS AMONG VERBAL SIZE-DISTANCE ASSOCIATION AND JUDGMENTS OF SIZE-CONSTANCY

($N=30$)

	Correlation	Size-constancy instruction		
		apparent size	objective size	perspective size
Verbal size-distance association	r_b*	0.20	0.44†	0.46‡
Apparent-size instruction	r§	—	0.39	0.56
Objective-size instruction	r	—	—	0.62

* Biserial correlation coefficient; verbal scores divided at median.

† $p<0.05$ (one-tailed probability), t-test between the two means used in computing r_b, $df=28$.

‡ $p<0.025$, same test as preceding one.

§ Product-moment coefficient; for one-tailed test, $p=0.01$ at $r=0.42$.

The skewed distribution of the verbal judgments of size-distance present some difficulty with respect to correlation. On the assumption that it would be possible to construct a scale for this variable which would yield normally distributed values, these scores were divided at the median, and biserial coefficients were computed for the correlations involving the verbal scores. These correlations and the intercorrelations among the scores for size-constancy are shown in Table I. The only nonsignificant relationship is that between the verbal scores and the size-constancy values obtained under the instructions for apparent-size.

Discussion. Undoubtedly the verbal associative scale used here could be considerably improved as a measuring instrument. It appears nevertheless, to have clearly reflected the existence of a cognitive, assumptive association between subjective notions of size and distance. In the authors' view, one demonstrates precisely the same cognitive association when he asks *O* whether objects 'look' smaller when they are farther away. There seems to be practically universal agreement that this proposition is true, at least among verbally competent members of Western culture. From this viewpoint the near-unanimity as to the direction of the association between the present word-items is not surprising. In any event, it is this conceptual kind of association between smaller size and greater distance on the part of *O* that we propose as the proper definition of the *perspective attitude*. This meaning appears to be in essential agreement with Gibson's use of the term, especially in the implication that the phenomenon depends as much upon characteristics of *O* as upon the particular visual stimulation presented in a given experiment.[4]

It was previously hypothesized that instructions regarding objective size in experiments on size-constancy lead to overconstancy because such instructions implicitly introduce the perspective attitude as the basis for judgment. The instructions for perspective size were deliberately designed to make the perspective attitude more explicit. Judgments of size-constancy under these two instructions should therefore show some correlation with an independent measure of the perspective attitude. It also was hypothesized, however, that the degree of deviation from size-constancy depends upon *O*'s motivational reaction to the situation; hence, a high correlation would be improbable. On the other hand, the instructions for apparent-size were specifically designed to eliminate the response-bias associated with the perspective attitude, at least for a group of *O*s as a whole, and should lead to no correlation with an independent manifestation of the perspective attitude. The correlations obtained in this study are consistent with these considerations.

The finding of a similar sort of verbal association between brightness and distance may have analogous implications for judgments of brightness under certain circumstances.

Summary. A purely verbal measure revealed a relationship between subjective notions of size and distance, and of brightness and distance. The verbal measure of size-distance was significantly correlated with size-constancy performance, depending upon instructions. It is suggested that a cognitive size-distance association defines the perspective attitude and leads to deviations from constancy in experiments on size-constancy.

---

[4] J. J. Gibson, The visual field and the visual world: A reply to Professor Boring, *Psychol. Rev.*, 59, 1952, 149-151.

APPARATUS

AN ULTRASONIC DEVICE FOR RECORDING ACTIVITY

By L. J. PEACOCK, University of Georgia, and
MARSHALL WILLIAMS, Athens, Georgia

Most of the methods of recording the activity of laboratory animals (such as revolving wheels, tambour- or spring-mounted cages, tilt-cages, and the like) involve some sort of mechanical coupling between the animal and the recording system which introduces an unknown amount of bias in the records. Efforts have recently been made to measure general activity photoelectrically or electromagnetically without mechanical linkages between the animal and the recording device.[1] Some of these methods have required elaborate instrumentation and few of them permit convenient adjustment of sensitivity to accommodate different experimental situations or different species of animals.

The electronic device reported here overcomes many of the disadvantages of the current methods of recording activity. It can be used with animals ranging in size from mouse to man and in any size enclosure from a standard mouse cage to a small room. Its sensitivity can be adjusted for measuring very small movements (such as respiratory motions of the thorax of an otherwise stationary rat) or only very large movements (such as locomotion). There is no mechanical coupling between the animal and the system, hence the measurement itself has no effect on activity. The electrical output of the device is amenable to counting, to inkwriting, and to storage on magnetic tape.

Functional description. The device is composed of an oscillator and an amplifier built on the same chassis and sharing a common power supply. The oscillator produces an ultrasonic signal which is converted into acoustic energy by a ceramic transducer. The ultrasonic signal radiates into the space containing the experimental animal (*S*), setting up nodes of standing waves. Internodal distance is determined by the frequency of the signal and by the geometry and nature of the reflecting surfaces enclosing the *S*'s space. The amplifier unit is connected to an identical

---

* The work described here was supported in part by research grant MY-3791(A) from the National Institute of Mental Health, US Public Health Service, and in part by the Graduate School of the University of Georgia.

[1] P. S. Siegel and J. A. Hard, An improved electronic device for the measurement of the activity-level of the small animal, *J. gen. Psychol.*, 42, 1950, 159-163; W. G. Mitchell, Differentiation of activity of three mouse strains with magnetic pickup apparatus, *Science*, 130, 1959, 455.

transducer mounted near the transmitter. As *S* moves within the field covered by the transducers, reinforcement and diminution of the signal detected by the receiving transducer occurs. These changes in signal caused by *S*'s movement are amplified, rectified, and emitted as pulses from the output tube of the device. The pulses can be used to drive a vacuum-tube, thyratron, or glow-tube counter circuit, or can be recorded on magnetic tape for automatic processing.

Circuit-description. A schematic diagram of the circuit of the apparatus is shown in Fig. 1. The ultrasonic oscillator (V5 and associated components) is a multivibrator, the frequency of operation of which is determined by a resonant circuit composed of a variable inductance (L2) and a fixed capacitance (C14). An oscillator frequency of 41 kilocycles has been used because of the commercial availability of transducers operating at this frequency. The variable inductance (L2) specified in the parts list is a television horizontal linearity coil adjustable from 0.2 to 3.0 millihenries; similar coils adjustable through a smaller or larger range could be used,

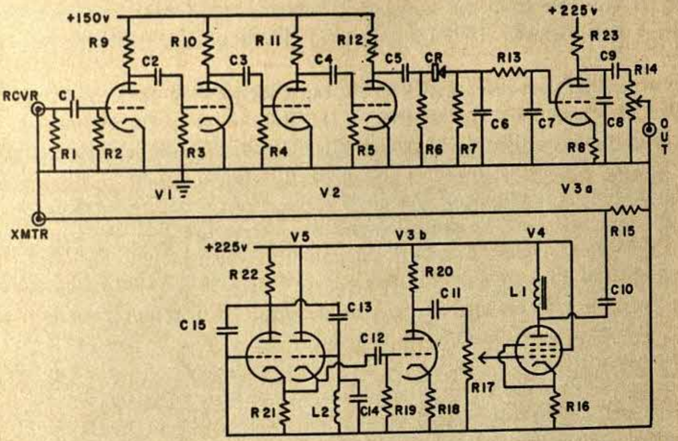


FIG. 1. SCHEMATIC DIAGRAM OF THE ULTRASONIC ACTIVITY-DEVICE

R 1-4.7 megohms;
R 2-R 6, R 9-R 12, R 22 = 100 kilohms;
R 8, R 18 = 1000 ohms;
R 13, R 15 = 1 megohm;
R 14 = 1 megohm potentiometer;
R 16 = 470 ohm, 1 watt;
R 17 = 500 kilohm potentiometer;
R 19 = 470 kilohms;
R 20 = 220 kilohms;
R 21 = 3.3 kilohms;
R 23 = 47 kilohms;
C 1—C 5, C 7, C 10 = 100 mmfd.;
C 6, C 11, C 12 = 0.005 mfd.;
C 8 = 0.01 mfd.;
C 9 = 0.1 mfd.;
C 13 = 0.005 mfd.; low-drift (Sprague PX-256 or equivalent);
C 14 = 0.01 mfd., low-drift (Sprague PX-116 or equivalent);
C 15 = 220 mmfd.;
L 1 = choke, 8 henries, 250 ohms (see text);
L 2 = horizontal linearity coil (Miller 6318 or equivalent);
CR = diode, Hughes type 1N198;
RCVR and XMTR = Westinghouse ceramic transducers, part number 753v 014H01;
V 1, V 2, V 3 = type 12AU7A;
V 4 = type 6AU6;
V 5 = type 12AT7.
All resistors are rated at 0.5 watts unless otherwise specified.

provided that inductances near 1.0 millihenry are obtainable for 41 kilocycle operation. To maintain a steady frequency of the oscillator, it is necessary to use low-drift capacitors in the tank circuit (C14) and for feedback (C13). The units specified in the parts list have proved to be stable.

The multivibrator is cathode-coupled to a buffer stage (V3b), which drives the ultrasonic output tube (V4). A small choke (L1) serves as the load for V4. The parts list specifies a filter choke for this purpose, but the primary of an audio output transformer from a small radio will work equally well, is smaller, and frequently is available from the junk-box. The ultrasonic output waveform is adjustable by a potentiometer (R17) in the grid circuit of V4, which serves as a sensitivity control.

The receiving transducer (RCVR) is capacitor-coupled to a four-stage voltage amplifier of conventional design (V1, V2). The output of the amplifier is applied to a crystal diode rectifier (CR) and a filter circuit. The diode specified, a Hughes type 1N198, gives the maximal response of several common diodes tried in this application. The output of the diode stage goes to the grid of V3a, which acts as a pulse-output tube. The potentiometer R14 serves to adjust the output level for various recording systems. If additional output is needed, R23 can be increased to 220,000 ohms, with some risk of clipping large pulses.

The two B+ voltages specified can be obtained from a conventional power-supply circuit with a capacitor-input RC filter.

Construction notes. Care in lead dress and component grouping is suggested. The tank-circuit leads should be short and rigid. A shield is recommended for V1. The power supply can be constructed on the same chassis (5 × 10 × 3 in.) as the oscillator and amplifier. The short shielded leads attached by the manufacturer to the transducers were not lengthened, since the apparatus is quiet in the audible range and can be placed next to *S*'s cage with no apparent deleterious effects. The transducers are components of a television remote control unit and are available from the manufacturer's parts distributors. Since the transducers are designed for 38-42 kilocycle operation, large departures from the suggested 41 kilocycle operating frequency will result in lower sensitivity.

Single-conductor shielded cables up to 25 ft. in length have been used to convey the output-pulses to recording equipment with very little signal loss.

Adjustment and calibration. Initial adjustment of the device requires a calibrated oscilloscope or the use of a signal generator with an uncalibrated oscilloscope (using Lissajous patterns for frequency determinations). After a 15-min. warm-up period, the oscillator frequency should be set by observing the ultrasonic output at XMTR on the oscilloscope while adjusting L2 to a recommended frequency of 41 kilocycles. When the oscillator frequency is correct, the wave-form should be adjusted by R17 to approximate a sine wave. Next, the pulse-output of the unit should be monitored on the oscilloscope at the output jack. With a 1-in. movement of the hand about one foot from the transducers, the pulse output should consist of a train of aperiodic pulses with a maximal peak-to-peak amplitude of 30 to 50 v. The output-pulse train can be adjusted as desired by resetting R17, and the output-level can be further attenuated by R14 to match the input characteristics of the recording device to be used.

Relative calibration may be accomplished by making the sensitivity adjustments while the unit is detecting *S*'s movements. Absolute calibration requires the use of a mechanical moving object, the extent of movement of which can be controlled. Calibration trials in this laboratory have utilized a chassis box moved by a shaft coupled to an eccentric cam driven by a controllable motor.

Installation. The transducers are mounted 3 or 4 in. apart, facing in the same direction. In one installation of three units currently in use, rats are suspended in individual cages (Fig. 2). Each cage and its ultrasonic device is contained within a plywood box lined with fiberboard. In another installation, the migratory unrest of groups of birds is recorded in entirely closed, fan-ventilated boxes. In this situation, the transducers are located in the ceiling of a box containing a cage of from two to six birds.

Installation of the device in spaces larger than the *Ss*' enclosures requires trial-and-error determination of the optimal location and orientation of the transducers. In certain applications in rooms, it appears to be advantageous to replace V1 or V2 (or both) with a tube giving greater gain, a type 12AX7. This is a direct substitution, with no wiring changes necessary.

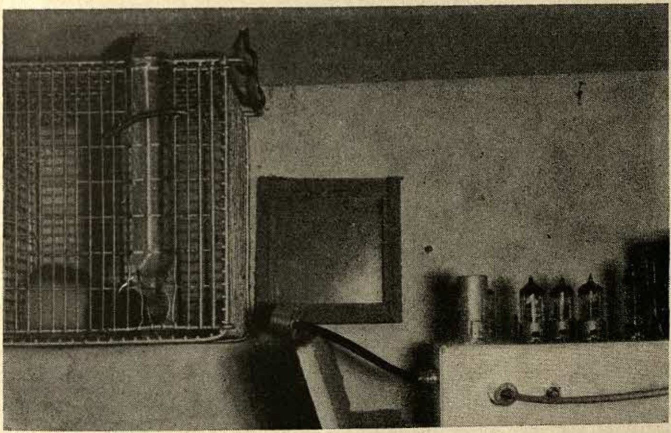


FIG. 2. INSTALLATION OF THE ULTRASONIC ACTIVITY DEVICE FOR RECORDING FROM RAT

Readout methods. Several methods of readout have been used. The simplest method uses the pulse-output to fire a tube-controlled electromagnetic counter. Because of the large number of counts (near 100,000 each 24 hr. from a female hooded rat) the life of the counter is brief, and expensive high-speed counters are needed to follow the rapid pulse-trains. An alternative readout method has used a glow-tube counter as a buffer stage between the ultrasonic device and an electromagnetic counter. Cumulative recorders of the type used in operant conditioning have been employed in situations in which a continuous record of activity is needed.

In the experimental situations in which an ink-written record of the output pulse-train is desired, the output signal can be applied to the power amplifier of a pen-motor; similarly, photographic recording from an oscilloscope is feasible.

Newbury has proposed the counting of the number of time-units in which activity occurs as a valid measure of activity, and has presented a relay circuit for the recording of such data.[2] The ultrasonic device is compatible with Newbury's system, and offers the advantage of greater sensitivity than the mechanical switches he assumed for detecting movement.

Discussion. Since the operating frequency of 41 kilocycles is well above the upper auditory frequency threshold of most animals, *S* is not coupled to the recording apparatus in any detectable way. The upper frequency threshold of the rat has not been measured precisely, although Gould and Morgan[4] report that the rat can hear 40 kilocycles and suggest that its hearing may extend two actives above that point.[3] Gould and Morgan were unable to specify absolute power levels of the auditory stimuli they used. We have been unable to condition rats in a shuttlebox to a 41 kilocycle signal of the amplitude used in the ultrasonic activity device. No pinna reflexes or changes in respiration were noted after 100 paired presentations of the 41 kilocycle signal and a painful shock. The same rats reached a criterion of 5 successive avoidance responses within 50 trials when a 12-kilocycle tone was used as the conditional stimulus. Retesting with the 41 kilocycle signal gave no evidence of generalization, the rats receiving shock on each of 50 additional trials. It appears that the 41 kilocycle signal at the intensity used was not detected by the rat.

The apparatus described here can be converted to 100 kilocycle operation (or any other frequency desired) provided that transducers can be obtained for the higher frequencies. In principle, then, it can be used at frequencies above the upper limit for any given species.

---

[2] Edward Newbury, Automatic measurement of general activity in time-units, this JOURNAL, 69, 1956, 655-659.

[3] James Gould and C. T. Morgan, Auditory sensitivity in the rat, *J. comp. Psychol.*, 34, 1942, 321-329.

A STABILIZED IMAGE REQUIRING NO ATTACHMENTS TO THE EYE

By T. N. Cornsweet, University of California, Berkeley

During normal vision, the eyes are in constant motion, and this motion causes continuous shifts of the retinal image with respect to the retina. To investigate the effects of such movement, a method called the 'stopped' or 'stabilized image' has been employed in a number of laboratories. This method allows the eye to move normally, but prevents the movements from producing corresponding shifts of the retinal image across the retina.[1] The basic finding is this: when the retinal image of an object is stabilized, the object rapidly fades out and disappears.

The principal difficulty encountered in the use of the 'stabilized image' has been that it requires the observer to wear a tight-fitting contact lens. There is always the possibility that such a lens will slip with respect to the optics of the eye. Another, much less severe, difficulty is that the visual phenomena that result from stabilizing the image cannot be demonstrated to anyone who has not been fitted with a special lens.

Three methods of stabilizing the retinal images that do not require a contact lens have been described in recent years. One involves imaging a point-source at the center of rotation of the eye, and another, the formation of an image of the iris at the retina.[2] Both of these methods overcome the effects of rotational movements of the eye, but each is subject to very profound artifacts caused by unavoidable lateral movements of head and eye. A form of stabilized image not subject to these artifacts was described by Ratliff.[3] He pointed out that Heidinger's brushes, the pattern visible when a uniform field of polarized light is viewed, are stationary with respect to the retina, and quickly disappear unless the plane of polarization is continuously changed.

The present paper describes a simple optical device that presents *O* with a stabilized real image without attachments to his eye. He sees a patch of light upon which the gross anatomical features of his retina are superimposed. The retinal image of the patch of light is not stabilized, but the image of the anatomical features, for example blood-vessels, is perfectly stabilized for small movements of the eye. When the apparatus is properly aligned, the blood-vessels rapidly fade out and the patch of light looks homogeneous.

---

[1] R. W. Ditchburn and D. H. Fender, The stabilised retinal image, *Optica Acta,* 2, 1955, 128-133; T. N. Cornsweet, Determination of the stimuli for involuntary drifts and saccadic eye movements, *J. opt. Soc. Amer.,* 46, 1956, 987-993; M. B. Clowes and R. W. Ditchburn, An improved apparatus for producing a stabilised retinal image, *Optica Acta,* 6, 1959, 252-266.

[2] J. ten Doesschate, A new form of physiological nystagmus, *Ophthalmologica,* 127, 1954, 65-73; R. W. Ditchburn, D. H. Fender, Stella Mayne, and R. M. Pritchard. A stabilised retinal image of the iris, *Proc. Phys. Soc.,* B. LXIX, 1956, 1165.

[3] Floyd Ratliff, Stationary retinal image requiring no attachments to the eye, *J. opt. Soc. Amer.,* 48, 1958, 274-275.

Apparatus. The design of the apparatus involves a few simple principles of geometrical optics. First, for any positive lens (*e.g.* convex on both sides), there is a distance, called the focal length of the lens, such that rays of light coming from any point on an object or image placed that far away from the lens will be parallel when they emerge from the other side of the lens. Conversely, any set of parallel rays that enters such a lens will converge to a point at a distance on the other side of the lens that equals the focal length. Secondly, if an object is placed at a distance from a lens that is twice the focal length of the lens, an image of the object will be formed two focal lengths distant on the other side of the lens, and the image will be the same size as the object.

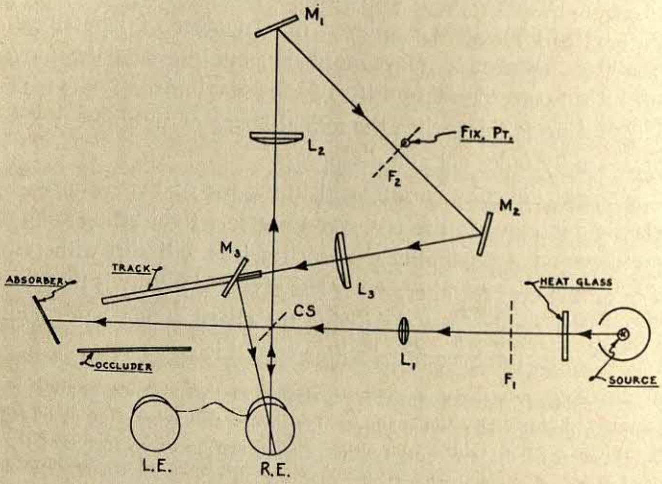


FIG. 1. SCHEMATIC DIAGRAM OF THE OPTICAL SYSTEM
Focal length of $L_1 = 147$ mm.; diameter $= 48$ mm.; focal length of L_2 and $L_3 = 220$ mm.; diameter $= 52$ mm. M_1, M_2, and M_3 are flat first surface mirrors.

Fig. 1 is a schematic diagram of the apparatus and *O*'s eye. Rays from a small source—*e.g.* tail-light bulb from an automobile—pass through heat-absorbing glass, through Lens L_1 and about 10% of the rays reflected from a cover-slip of a microscope (CS) into the eye. (The other 90% are trapped by a black absorber.) Lens L_1 is placed at twice its focal length from the source, so that it forms an image of the filament of unit-magnification in the plane of *O*'s pupil. As long as this image is smaller than the natural pupil (say less than 2 mm. in diameter), and is centered on the natural pupil, changes in the size of the pupil will not affect the intensity of light at the retina.

When the eye is accommodated for infinity, the optics of the eye will form an image on the retina of any object whose rays are parallel when they enter the eye. Therefore, any object in the plane labeled F_1 (at the focal distance from L_1) will form a sharp image on the retina. In the present apparatus, it is convenient to place a square aperture in that plane, so that a bright square of light is formed on the retina. *O* is instructed to fixate a point that appears off to the left of the incoming beam. Therefore, the square of light is formed on his peripheral retina.

A very small fraction of the light striking the retina is reflected back out of the eye along the same general direction that it entered. About 90% of those rays pass through the cover-slip, travel around the path through lenses L_2 and L_3, and eventually head back toward the eye. The rays from the square will be parallel as they emerge from the eye. Therefore, an image of the square and of the (peripheral) retina under the square will be formed at F_2, the focal plane of lens L_2. The light then continues from Mirror M_2 through Lens L_3, which is at its focal distance from F_2. Therefore, the rays from the retina are again parallel as they reflect from Mirror M_3 and enter the eye. The optics of the eye then reform the image of the square and peripheral retina on the central retina.[4] A very small, dimly lit bulb in the Plane F_2, a little off the axis of the system, serves as a reference for fixation during the alignment of the optics.

As long as the focal lengths of lenses L_2 and L_3 are equal, the image of any peripheral blood-vessel that is formed on the central retina will be the same size as the blood-vessel itself. In other words, the system has a magnification of ×1.0. For this reason, any rotation of the eye will produce an exactly equal shift of the retinal image of the blood-vessel, and as long as there is an odd number of mirrors in the viewing path (in this case three), the shift of the retinal image will be in the same direction as the shift of the retina itself. The image of the blood-vessels is thus stabilized against rotational and torsional eye-movements. It is also true in this system that the image of the retinal blood-vessels is stabilized against translational movements of the head and eyeball. The rays from a blood-vessel are parallel as they leave the eye. Lateral movements of the eye cause the retina and the refractive surfaces of the eye to move equal amounts (by definition). When the rays are parallel, such a movement will produce no shift of the position of the image of the blood-vessel that is formed in the Plane F_2. Conversely, since the rays are parallel as they reënter the eye, lateral shifts of the eye produce no shift of the retinal image with respect to the retina (the image and the retina move through the same distance in the same direction). For these reasons, the optical system produces an image of the retinal structures that is stabilized with respect to all possible movements of the eye: rotational torsional, and translational. The image of the square of light is also visible, but that image is not stabilized against rotational or torsional movements.

As the system has so far been explained, the Lenses L_2 and L_3 are not necessary. The logic above would apply equally well to a system of mirrors alone (except that there must then be an even number of mirrors). If the eye is placed in the correct position in the optical system, as in Fig. 1, but Lenses L_2 and L_3 are simply removed, the image formed on the central retina is of both the peripheral retina and of the pupil of the eye. One is in sharp focus and the other is a little blurred. In this case, it is possible to see only a very small part of the peripheral retina through the small pupil-aperature. Lenses L_2 and L_3 eliminates this problem in the following way. L_2 is approximately at its focal distance from the pupil, so the rays taken to originate from each point on the pupil are parallel when they strike Lens L_3. Lens L_3 is also approximately at its focal distance from the pupil, so that it forms an image of the pupil at the pupil. Thus the rays from the pupil are altogether out of focus at the retina, and the pupil no longer limits the size of the visible region of retina.[5]

---

[4] If *O* needs glasses for far vision, he must wear them in this apparatus.

The particular lenses used in this apparatus are not critical, as long as the focal lengths of L_2 and L_3 are equal. The lenses listed in the legend of Fig. 1 are inexpensive. Mirrors M_1 and M_2 should be first-surface mirrors, to avoid doubling the image. Any thin piece of flat glass will do as the cover-slip (*CS*) in Fig. 1.

Adjustment and alignment. The initial alignment involves so many factors that it is rather difficult to accomplish and just as hard to explain. *O*'s chin should be supported in a rest that can be easily adjusted up and down as well as laterally. It is probably easiest to do the first stages of aligning without Lenses L_2 and L_3. When the mirrors are properly positioned, the incoming beam will produce a glare in peripheral vision and, through the viewing path, the viewing eye will be visible, the pupil appearing to be lit with a pink light. The Lenses L_2 and L_3 may then be inserted, and, after some relatively small readjustment of the mirrors, the pink light from the retina will again be visible, but this time the outlines of the pupil will not. Instead, a pink figure the shape of the aperture in Plane F_1 will be visible. This figure is actually the retinal image of the aperture. If the alignment happens to be such that this image falls on a blood-vessel, the vessel will be clearly visible at the same time. If no vessel is visible, one may be brought into view by sliding Mirror M_3 along its track and rotating it appropriately at the same time, thus shifting the bright image across the retina.

When the eye and optical system are perfectly aligned, *O*'s view of his retina will be obscured by a diffuse white haze or a blob of white light, the light reflected directly from the cornea. Moving the head very slightly to one side will misalign the system enough to lose the corneal reflection while retaining the light from the retina itself.

When *O* is properly positioned and the optics correctly adjusted, the retinal structures will disappear so rapidly that the system seems to be going badly out of focus. Eye-movements beyond the field of Lens L_3 and back will restore vision temporarily. Because proper alignment results in poor seeing, it is sometimes easier, during the early stages, to align the system by viewing the retina of the right eye with the left eye. To do this, the occluder is removed, Mirror M_3 is moved along its track until it is roughly in front of the left eye, and it is then rotated until the light passing through L_3 is directed toward the left eye. When this is properly done, it is very easy to see the retinal blood-vessels of the right eye through the left one; and, because the two eyes do not move together in perfect synchrony, the blood-vessels will not disappear.[6]

---

[5] This is a verbalization about a fact that can be derived unambiguously from the principles of geometrical optics.

[6] John Krauskopf, T. N. Cornsweet, and L. A. Riggs, Analysis of eye movements during monocular and binocular fixation, *J. opt. Soc. Amer.*, 50, 1960, 572-578.

ELECTRICAL AND ELECTRONIC DEVICES FOR THE TEACHING OF SAMPLING STATISTICS

By ALBERT E. BARTZ, University of Arizona

During the first course in statistics, one of the more difficult concepts for the beginning student to grasp is that of the *standard error* and its relation to the sampling distribution. This may be due, at least in part, to the difficulty of relating a sampling distribution to physical data. Many teachers of statistics have overcome this difficulty by constructing a population of numbered tags, beans, beads, etc., with a known mean and standard deviation. The student then extracts samples of a certain size, constructs a frequency-distribution of the sample means, and calculates the standard deviation of this distribution. He has now *calculated* the standard error, and become familiar with the concept of the sampling distribution and the standard error. Attesting to the success of this procedure is its wide use by many teachers of statistics, and its inclusion in many texts on statistics. The main disadvantages of this procedure are the time and effort required to select a number of samples. The process is time-consuming, especially when a large number of samples is to be drawn.

The present research was undertaken to develop both an electrical and an electronic device that would perform this function quickly and accurately.[1] These devices were designed to sample randomly from a designated population and to present the sample to the operator.

Electro-mechanical device. This apparatus employs a motor-driven arm that rotates a sensing unit over a circular pattern of contacts divided into sectors (Fig. 1). A population-distribution is 'built in' by arbitrarily choosing score values of 1 to 13. A normal distribution is constructed with these score-values, and the length of each sector representing a score is determined by the area under the normal curve. The mean of this distribution is 7 with a σ of 2.0. The length of each sector, is degrees of arc, with the corresponding area under the normal curve is given in Table I. There is a separation of 1° between segments for purposes of insulation. In addition to the normal distribution, a skewed distribution was also constructed. A selector-switch on the front panel sampling from either distribution.

When the device is turned on, the arm rotates at 1 r.p. sec. To take a 'reading,' the operator presses a spring-return switch, the arm coasts to a stop, and the sensing unit makes contact with one of the 13 sectors. When this contact is made, the appropriate reading is presented to the operator by one of the 13 indicator-lights on the front panel. As soon as an indicator-light comes on, the observer releases the switch, writes down the observation, and the arm starts moving again. The randomness in taking observations is provided by the time required to enter each reading. Small variations in this interval interrupt the periodicity which might occur due to the relatively constant rate of movement of the arm.

---

[1] The construction of this device was made possible by Research Grant G-12353 from the National Science Foundation. The author is grateful to Dr. Robert W. Heath, Dr. Douglas J. Hamilton, and Mr. Robert Lucas for their advice and assistance in the design and construction of the apparatus.

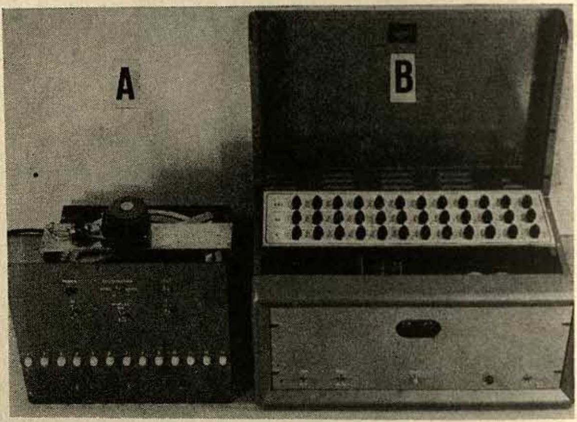


FIG. 1. THE SAMPLING DEVICES
A, the electro-mechanical unit (cabinet removed); B, the electronic unit (cover raised to show selector-switches).

TABLE I

CONSTRUCTION OF SECTORS AND RESULTS OF 2600 TRIALS

Score-value	Portion of area under normal curve	Length of sector	Observed (2600 cases)	Expected (2600 cases)
1	.002	0° 41′	11	5.2
2	.009	3° 7′	25	23.4
3	.028	9° 43′	61	72.8
4	.066	22° 54′	163	171.6
5	.121	42° 0′	311	314.6
6	.175	60° 44′	442	455.0
7	.198	68 42′	544	514.8
8	.175	60° 44′	437	455.0
9	.121	42° 0′	329	314.6
10	.066	22° 54′	179	171.6
11	.028	9° 43′	63	72.8
12	.009	3° 7′	25	23.4
13	.002	0° 41′	10	5.2
	1.000	347°	2600	2600

Two other features of this device should also be noted. Due to the low probability of observations occurring at the extreme ends of the distribution, a test-switch is included that indicates when an indicator-light is not functioning. After any given number of observations, the operator may test the indicator lights to verify that the non-appearance of extreme values is not due to a malfunctioning light. Also included is a remote cable that enables the instructor to demonstrate the sampling at some distance from the device.

Electronic device. It will be convenient to describe the operation of the electronic device by comparing it with the drawing of numbered tags from an artificial population. In the latter case, one chooses a set of numbers and determines the distribution of the set by assigning a certain amount of tags to each number of the set. For the electronic device, the set consists of the numbers 1 through 12, which are represented by the 12 positions of Counter B (see block-diagram in Fig. 2). A panel is provided with three selector-switches for each of the 12 numbers. These switches are used to select the number of tags that is desired for each of the 12 numbers; the three switches for each number represent units, tens, and hundreds, so that up to 999 tags may be selected for each number. The 'tags' themselves are represented by electrical pulses from the pulse-generator.

To use the device, Switch S is depressed for several seconds. When the switch is released, the position of Counter B at that instant is displayed by the indicator. The probability of a particular number's appearing in Counter B is dependent upon the number of counts (*i.e.* 'tags') which has been set in the selector-switches for the particular number. Thus depressing and releasing S corresponds to withdrawing a tag from the container. The number displayed may be recorded and a distribution may then be plotted.

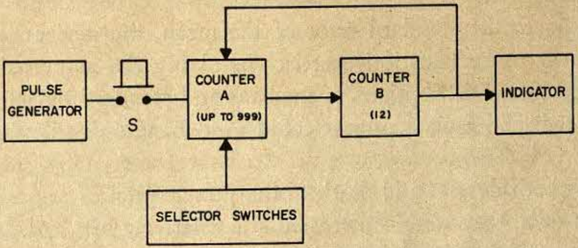


FIG. 2. BLOCK DESIGN OF THE ELECTRONIC UNIT

The indicator is a neon display of Arabic numerals easily visible from a distance of 30 ft. The display remains fixed until Switch S is depressed, when it is extinguished until S is released.

An example will illustrate the operation of the device. It is assumed that, prior to Switch S' being depressed, Counter B is in the one position and Counter A is reset to 000. Suppose that the desired distribution requires that the number one have 421 'tags'; the number two, 380 'tags'; etc. The selector switches are set accordingly. Now when Switch S is depressed, the pulse-generator is connected to Counter A, which begins to count. When Counter A reaches 421, it is reset to 000 and Counter B is advanced to 2. Counter A continues to count pulses and when it reaches 380, it is again reset to 000 and B is advanced to 3. This procedure continues until Counter B reaches 12; after the required number of pulses has been counted for the number 12, Counter B is advanced to 1 and the cycle begins again. At any time during the cycle that Switch S is released, the number in Couner B is displayed, as previously described, and the number in Counter A remains there. When Switch S is again depressed, the cycle resumes at the point at which it was interrupted. The entire cycle is completed in less than 1 sec. Any desired distribution may be chosen merely by setting the proper numbers in the selector-switches.

Reliability and classroom use. Table I shows the results of 2600 observations taken in 104 samples of 25 cases each with the electromechanical device. A Chi-square test between these and expected observations yields a value that is not significant. Also, the standard deviation of the distribution of sample means is 0.36, while the value of $\bar{\sigma}_m$ is 0.40. These results indicate that the device is serving its purpose with accuracy.

At this writing the electronic device was undergoing extensive classroom use. Results obtained in preliminary tests indicate high reliability in sampling not only from a normal distribution, but from other shapes of distributions as well.

In addition to the standard error of the mean, the devices lend themselves to many other statistical exercises involving the nature of the sampling distribution. Distributions of medians and standard deviations could be constructed, and a class project could even handle distributions of the Pearson *r*.

Although both devices add much to the presentation of sampling theory in the classroom, they were constructed at a relatively low cost. Because of its simplicity, the electro-mechanical device was relatively easy to construct, and could be assembled in any workshop. The electronic sampling unit required more work but could be constructed by an electronic technician.[2]

---

[2] Wiring-diagrams and details of construction may be obtained from the author, now at Concordia College, Moorhead, Minnesota.

THE PSYCHOLOGICAL LABORATORIES AT WAKE FOREST COLLEGE

By Robert H. Dufort, Wake Forest College

Winston Hall, the new life sciences building at Wake Forest College, was dedicated on November 3–4, 1961. The building, first opened in August, houses the Departments of Psychology and Biology. This note describes the building and the space occupied by the Psychology Department.

The building (252 × 100 ft.) has three floors arranged in the plan of a modified H. The front entrance is located on the first floor, while rear entrances are located on both the ground and first floors. General class and lecture areas are in the center section of the first floor, on either side of a lobby and display area. Advanced classrooms and laboratories, as well as faculty research areas, are located in the wings of the ground and second floors, away from traffic noise. Heating and air-conditioning are combined in a single system; thermostatically operated vents control the proper mixture of hot and cold air to provide the precise temperature desired in the various rooms of the building. Space is divided into three categories: psychology rooms, biology rooms, and rooms used jointly by both departments. In the sections which follow, the floor plans of the building are given, and the rooms falling in the first and third categories are described.

Ground floor. Fig. 1 presents the plan of the ground floor. Psychology uses 10 rooms on this floor.

Joint rooms. Starting in the center section of the building, room 27B (24 × 15 ft.) is a shop for wood and metal work. The shop is located under the rear steps from the first floor to the ground to reduce the spread of noise to other areas. Room 29 contains the building incinerator. Room 24 is a statistics laboratory furnished with work tables that are equipped with electrical outlets for the operation of calculators. Both student and research calculators are used and kept in an adjoining storage area. Rooms 16 and 18 are used, respectively, for storage of the supplies for animal colony and for animal surgery. These rooms service the main animal colony in Room 6 (25.5 × 19.5 ft.).

Psychology rooms. The suite of rooms that includes 2, 2A, and 4 is the undergraduate experimental psychological laboratory. Rooms 2 and 2A are equipment storage and preparation areas that can be converted readily to individual, research laboratories when they are not being used to service the undergraduate course. Room 4 contains 15 research cubicles, each about 6.5 × 6 ft., that open into a central lecture and demonstration area (31 × 16.5 ft.). Each cubicle is designed to be used by a pair of students and is furnished with a table, two chairs, and a wall cabinet for equipment.

A photography darkroom in Room 20 completes the description of the ground floor.

First floor. Fig. 2 presents the plan of the first floor. Psychology does not have exclusive use of any rooms on this floor; it has, however, the joint use of three lecture rooms; A, B, and C.

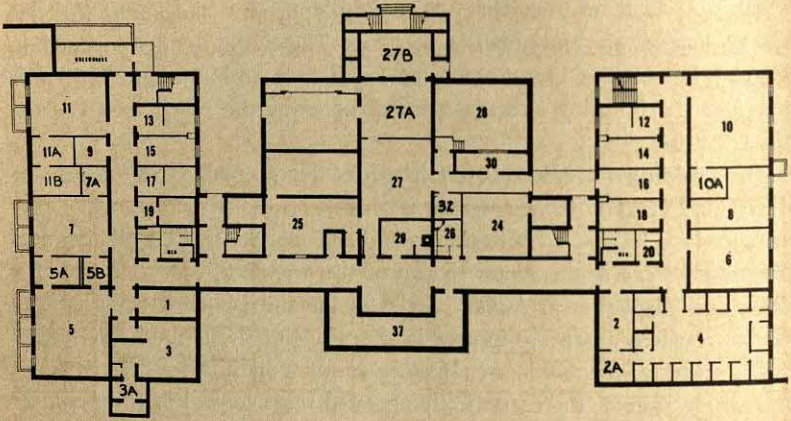


FIG. 1. GROUND FLOOR PLAN
(Front of building is at the bottom.)

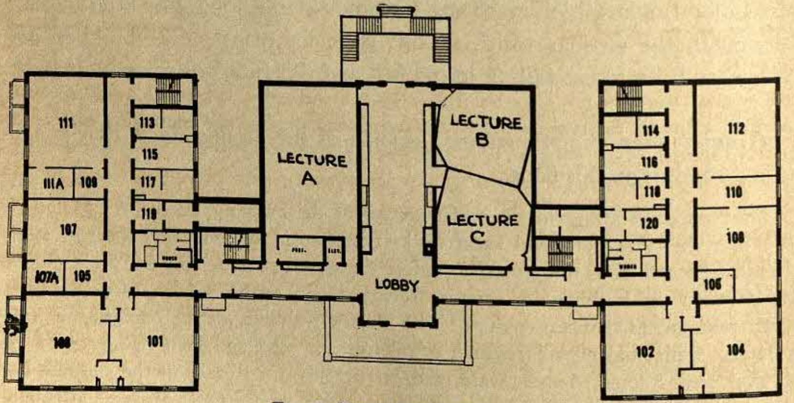


FIG. 2. FIRST FLOOR PLAN

Joint rooms. Room A is a small auditorium with a seating capacity of 167. Upholstered theater seats with folding desk arms are so arranged in rising tiers that the room actually occupies both the first and ground floors. A projection-booth is located near the first floor entrance. Switches in the front of the room give a lecturer remote control over lighting in the room, projection-booth equipment, and a battery of blackboards and screens that can be brought before an audience on sliding tracks from an adjoining preparation area (Room 27A on ground floor).

Rooms B and C are identical in size and have a seating capacity of 70 each. The horn shape of these rooms, and the arrangement of seats in rising tiers, combine to produce excellent sight lines and acoustics.

Second floor. The psychology offices and general research laboratories are located on this floor. A total of 25 rooms is used. Fig. 3 shows the floor plan.

Joint rooms. Starting again in the center of the building, Rooms 221 and 223 are classrooms for advanced, non-laboratory courses, with seating capacities of 30 and 48 respectively. Rooms 234 and 236 are a conference room and a lounge. A small kitchen connects these areas. Room 226A contains duplicating equipment and secretarial supplies.

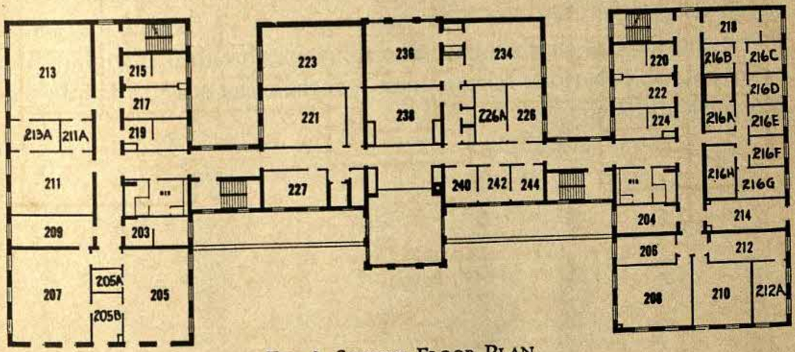


FIG. 3. SECOND FLOOR PLAN

Psychology rooms. The offices of the department chairman and secretary are located in Rooms 240 and 242. Rooms 204 and 206 are research laboratories. The large corner area, Room 208 (24.5 × 19.5 ft.), is an advanced undergraduate laboratory currently used for instruction in physiological psychology. Rooms 210 and 212A are an observation suite, separated by a one-way viewing window. Observation can be made from either room into the other, depending on the lighting arrangement. A two-way intercommunicational system joins the rooms. Entry to 212A is made through Room 212, which functions individually, or in connection with the observational suite, as a research laboratory. Room 214 serves as a seminar and reading area; shelves along one wall hold current issues of the periodical literature.

The remainder of the psychology space is entered from an inner corridor set back from the main hallway. This arrangement has resulted in the virtual elimination of noise and traffic from this area. Room 216A is a control area for an adjoining soundproof research room. The soundproof room (7.5 × 7 ft.) has its own ventilation system and is electrically shielded. Room 216B is a research laboratory. An electronics shop is located in 216C. Rooms 216D, 216E, 216F, and 216G are faculty offices. Room 216H is a partially soundproof research laboratory. Finally, a colony for animals used in research on this floor is maintained in the two rooms at the end of the corridor (218).

Several additional features of the building are of general interest to psychologists. The Biology Department maintains a radiational laboratory with facilities for isotope storage. Biology also uses four constant-temperature rooms which provide a range from slightly below room-temperature to well below freezing. A television studio furnishes facilities for closed-circuit viewing in selected areas of the building.

Winston Hall offers unusually fine opportunities for teaching and research, as well as for interdisciplinary work in biological psychology. For many years, psychology was taught at Wake Forest College by a combined Department of Psychology and Philosophy. An independent department was established in 1958. The rapidity with which the department has received its new quarters has been encouraging to psychologists at Wake Forest and to other psychologists who recognize that their discipline, as a biological and behavioral science, must join the older sciences in adequate laboratory surroundings.

NOTES AND DISCUSSIONS

A NOTE ON INCREASING THE EFFICIENCY OF LUCHINS' MENTAL SETS

The water-jar problems employed by Luchins have been widely used as classroom demonstrations and as problems for research.[1] The senior author has used them for many years and was frequently frustrated by the failure of students to acquire the appropriate sets due to errors (apparently) in computation. An examination of Luchins' problems suggests that the formation of sets may be hampered by the relative complexity of the problems. The solution normally expected follows the formula B–A–C–C (see Table I) but the sizes of the jars as well as the position of the larger (A or C) vary and frequently determine what the student (*S*) does. *S* may change from a B–A–C–C operation to a B–C–C–A operation. Should *S* make a mistake in subtraction, he may reject the set-type of solution.

TABLE I

Luchins' Original Problems and Their Revision

Problem	Original problems				Proposed revision			
	jars given			to get	jars given			to get
	A	B	C		A	B	C	
(1)	29	3		20	43	89	2	42
(2)	21	127	3	100	25	59	2	30
(3)	14	163	25	99	32	69	3	31
(4)	18	43	10	5	52	78	3	20
(5)	9	42	6	21	43	93	4	42
(6)	29	59	4	31	31	61	4	22
(7)*	23	49	3	20	17	37	3	14
(8)*	25	39	3	18	41	86	4	37
(9)†	28	76	3	25	47	68	4	51
(10)*	18	48	4	22	27	59	5	22
(11)*	14	36	8	6	13	35	3	16

* Check problems, solvable by other methods.

† Extinction problem, not solvable by set method.

To avoid these difficulties, the new set of problems shown in Table I was constructed in which all of the subtractions were simple (no 'borrowing') and A was always larger than C. The A and B jars were described in two-digit numbers, the C jar was always a one-digit number. Thus, *S* was encouraged to attack the problems in a systematic and routine manner both by the order of events and the ease of the computations.

---

[1] A. S. Luchins, Classroom experiments in mental set, this Journal, 59, 1946, 295-298.

When 18 *S*s (college students) were tested with the new problems, they acquired the expected set in 2.65 problems as determined by subjective reports of 'insight.' When the time of solution was plotted, a reliably (two judges) determinable drop in solution-time was revealed after 2.85 trials in contrast to a similar drop after 4.58 problems for 17 *S*s with Luchins' original problems.

With the discovery of the relatively rapid formation of the sets, it also became apparent from *S*s' reports that they frequently "caught on" after the first problem, tried to solve the next in the same way, and used the third and perhaps the fourth problem to check the reliability of the method. Thereafter they continued to use it because it worked and no alternative procedures seemed called-for. This left open the question of the role of frequency or practice in strengthening the set. Does additional practice actually strengthen a set, as implied by Luchins, Mayzer, Tressell and Leeds, and Youtz?[2] Luchins had already suggested a test in the 'extinction' problem wherein a problem is offered which cannot be solved by the set-method. Subsequently *S*s are again offered a chance to use the set-method or some alternate procedure. It was decided to test the hypothesis that additional practice in following the set-method would result in more solutions of the set-type after an 'extinction' problem was attempted.

TABLE II

PERCENTAGE OF *S*S USING THE SET SOLUTION IN THE CHECK- AND EXTINCTION-TRIALS

Group	Number of training trials	First check trial	Extinction trial	Second check trial
1	3	100	26	74
2	6	100	33	60
3	9	100	33	74
4	12	100	46	60

Four groups of 15 *S*s were provided with specially prepared booklets of problems (randomly arranged across groups) and were given different amounts of practice (3, 6, 9, or 12) trials in solving problems in the set-fashion.[3] Following the practice they were given a 'check' problem which could be solved by following the set or by an A + B or A − C procedure. An 'extinction' problem followed and then another 'check' problem appeared. The results appear in Table II, from which it can be seen that all four groups solved the first check problem in accordance with the set, testifying to the effectiveness of the set even after three trials. There is a steady but not statistically significant rise in the failures on the extinction-problem (no solution in 2 min.), but no reliable differences in the percentages of solutions of the second check problem. The *S*s with 12 practice trials produced no more set-solutions than those with 6 trials. In actual percentages they showed even less tendency to use the set-procedure than the 3-problem group. The findings suggest that the Luchins' type of set can be quickly acquired (if the problems are simplified), tested for effectiveness, and then used indiscriminately even after a few successful tests. Additional confirmation trials do not appear to strengthen the set once it fails to be effective.

---

[2] Luchins, Mechanization in problem solving: The effect of einstellung, *Psychol. Monogr.*, 54, 1942 (No. 248), 1-95; M. S. Mayzer, The effects of competition of various strengths of sets in problem solution, *Psychol. Newsletter*, N.Y.U., 6, 1955, 134-155; M. E. Tresselt, and D. S. Leeds, The einstellung effect in immediate and delayed problem solving, *J. gen. Psychol.*, 49, 1953, 87-95; R. E. Youtz, The relation between the number of confirmations of one hypothesis and the speed of accepting a new and incompatible hypothesis, *Amer. Psychologist*, 3, 1948, 248-249.

University of Buffalo

B. R. BUGELSKI
E. M. HUFF

PERCEPTUAL UPRIGHTNESS AND COMPLEXITY OF RANDOM SHAPES

The following experiments may be of interest to investigators who are using random or nonsense-shapes as visual stimuli in their research.[1] Results of the first of these experiments indicate that, for many *O*s, a random shape has a preferred position. Data from the second of these experiments show that reliable judgments of complexity can be obtained when all the random shapes have the same number of 'points,' a physical characteristic of nonsense-shapes which has been shown to be highly related to judgments of complexity.

Twenty closed-contour, random shapes, constructed according to Method 1 described by Attneave and Arnoult, were presented to 30 men and women college students.[2] Each shape was cut out of black paper and separately mounted on a 9-in. white cardboard disk. During the experiment, the cardboard disks were attached with double-faced scotch tape to a metal disk which was free to rotate around a horizontal axis. Each *O* was instructed to "turn each shape around (360°) once and look for a position in which the shape appears right-side-up." The shape could then be rotated to its "right-side-up" position. Readings of *O*'s settings were taken from a 360° protractor on back of the disk. It was found that for every shape there was one position that was judged to be upright by at least 30% of the *O*s. For each of 10 shapes, one position was judged upright by 50–83% of the *O*s, and of the remaining 10 shapes, 5 were judged upright in *two* positions by a combined total of 69–83% of the *O*s. Considering the large number of possible positions which could have been selected, there is little doubt that the small number actually obtained indicates that random shapes have 'upright' positions.

---

[3] Providing problems in booklet form, one to a page, proved significantly more effective in another study than displaying them all on one page.

[1] Fred Attneave, Physical determinants of the judged complexity of shapes, *J. exp. Psychol.*, 53, 1957, 221-227; L. L. Elliott, Reliability of judgments of figural complexity, *ibid.*, 56, 1958, 335-338; A. G. Goldstein, Familiarity and apparent complexity of random shapes, *ibid.*, 62, 1962, 594-597.

[2] Fred Attneave and M. D. Arnoult, The quantitative study of shape and pattern perception, *Psychol. Bull.*, 53, 1956, 452-471.

The judgments seemed to be based on one or more of the following 'plans': (a) placing one of the longer, straight sides of the shape parallel to the horizon; (b) placing the ends of two projecting points on the visual horizon; and (c) balancing the shape on a point or side.

In the second experiment, 30 *O*s judged the complexity of 40 random shapes (constructed in same manner as previous experiment) using a 7-category rating-scale.[3] Each shape had 12 inflection-points, or turns on its contour. The sign-test was applied to the data and mean ratings as small as 0.5 differed significantly at or below the 1% level, indicating a high degree of reliability of judgment. The shapes judged least and most complex had mean complexity-ratings of 0.97 ($SD = 0.99$) and 4.86 ($SD = 0.93$), respectively. There ratings compare favorably with the values obtained in another experiment where 4-, 6-, 8-, 12-, 16-, and 24-point figures were judged; mean rating for all 4-point shapes was 0.97 and for all 24- point shapes, 5.70.[4] Obviously, the *O*s were making their judgments in the present experiment on the basis of information other than the number of turns. Although no systematic attempt was made in this pilot study to determine what these variables were, angular variablility (*i.e.* jaggedness of the figure) and compactness (perimenter/area), two of the parameters investigated by Attneave,[5] were clearly influencing the *O*s' judgments. When the figures were ordered according to judged complexity, those with high mean ratings were characterized by acute angles, lack of a main 'body,' and relatively long perimeters in relation to a very small enclosed area, whereas the opposite was true for those shapes with low mean ratings. In Attneave's study of figural complexity, the effect upon judgments of any factor other than number of turns was minimal and relatively difficult to detect by statistical procedures. Because of the impressive nature of the main variable, number of turns, perhaps his *O*s' judgments were almost not affected by the other stimulus-variables. By employing the method described in the present experiment it is possible to evaluate directly the effects upon judgments of any number of shape-variables. To keep one variable constant while investigating the effects of changing a second variable is, of course, not a new method in psychology, but this experiment demonstrates that, in form-perception, making all stimuli identical on one parameter does not make an impossible task for the judges.

---

[3] Categories of the scale were the same as those used by Attneave, *op. cit.*, 223: 0, extremely simple; 1, very simple; 2, simple; 3, medium; 4, complex; 5, very complex; 6, extremely complex.

[4] Goldstein, *op. cit.*, 595.

[5] Attneave, *op. cit.*, 222-223.

University of Missouri

ALVIN G. GOLDSTEIN
JAMES ANDREWS

ON THE RELATIONSHIP BETWEEN ESTIMATES OF MAGNITUDE OF LOUDNESS AND PITCH*

In the course of his interesting attempts to develop a model of choice behavior, Luce has turned his attention to the derivation of scales of subjective magnitude, and to accounting for the "power law."[1] Partly because of the upsetting fact that loudness exponents derived from his model are at least an order of magnitude too large, he turns to a consideration of frequency as the joint determiner with intensity of a ("cupped") psychological surface defined by the two. Making his assumptions, it follows that: "Over a population of subjects we also predict an inverse correlation between the intensity exponent and the range of numerical values assigned to frequency."[2] An opposite view may be derived from the results of Jones and Marcus, which suggest that the range of numbers used by a given *O* may reflect merely a response bias having a little more meaning than simply referring to *O*'s past experience.[3] It is apparent, therefore, that these two points of view make opposite predictions about judgments of loudness and pitch. Fortunately, the question of which prediction is correct is easy to decide, and the experiment to be reported here is aimed at making that decision.

---

* This study was supported by a grant (G-14260) from the National Science Foundation.

[1] R. Duncan Luce, *Individual Choice Behavior*, 1959, 38-54. For the "power law" see S. S. Stevens, On the psychophysical law, *Psychol. Rev.*, 64, 1957, 153-181.

[2] Luce, *op. cit.*, 54.

We have obtained judgments of loudness and of pitch from 21 *O*s, all students in a course in elementary psychology. Judgments of loudness were obtained for 13 1000-cycle tones ranging, by 5-db. steps, from 40 to 100 db. The stimuli were presented in a balanced random order until six judgments had been obtained for each level. Instructions were of the usual magnitude-estimate variety, the standard stimulus being set at 70 db. and called 10.[4] Pitch was judged in the same way for 13 different frequencies, ranging from 50 to 7900 cycles per sec. The different frequencies were made subjectively as nearly equally loud as conveniently possible by judgments of two trained *O*s when presented through the earphones (Telephonics TDH 39) used in the experiment. Ten *O*s judged pitch at a first session, loudness a week later, while the other 11 reversed this sequence. No one noted any significant differences in loudness among the different frequencies.

Although a simple comparison of the ranges of numbers used by the *O*s for the two dimensions would probably be adequate, it was decided to meet the letter of Luce's prediction. Therefore, curves of the form $\Psi = aI^b$ were fitted by least squares to the geometric means of each *O*'s estimates of loudness. The resulting exponent, b, was compared with the range of numbers the given *O* had assigned to the frequencies. The product moment correlation between these two values is +0.44, which, for $N = 21$, is significant at the 5% level.

On the basis of the results, it may be concluded that Luce's prediction has to be rejected. Further than this, the alternative suggestion of a positive, not negative or zero, correlation is upheld at the 5% level of significance. These results suggest that the Luce model underlying his discussion of interaction of continua is in error, since the model leads to the incorrect prediction. The difficulty is not, of course, with the mathematics. The difficulty most likely lies in the assumption of separate 'storage' of loudness and pitch scales. This assumption implies that our retrieval operations are also independent, so that a tendency to emit large numbers in one situation is independent of the tendency to emit large numbers in another situation. It is not, in other words, possible to separate the 'stored' scales from the procedure used in constructing them, and the procedure, magnitude estimation, is the same in both cass. This being the case, it is not astonishing that response-biases are carried from one situation to the other. The same kind of bias probably underlies the "personal exponent" of Jones and Marcus.[5] This bias may also account for the tendency for shorter intensity-ranges to result in steeper functions of subjective magnitude. This is true both as among different senses with different intensity-ranges and also in regard to restriction or expansion of range within a given modality.[6] A given *O* has, then, a characteristic response-range which he tends to expand or contract to cover the range of stimuli presented. No more complicated a theory than this is needed to account for the positive correlation we have found.

---

[3] F. N. Jones and Maxine J. Marcus, The subject effect in judgments of subjective magnitude, *J. exp. Psychol.,* 61, 1961, 40-44.

[4] For the most complete discussion of the method of magnitude estimation see S. S. Stevens, The direct estimation of sensory magnitudes-loudness, this JOURNAL, 69, 1956, 1-25.

University of California
Los Angeles, California

F. NOWELL JONES
MORRIS H. WOSKOW

A NOTE ON RECENCY AND PERCEPTION

Epstein and Rock recently reported a group of experiments employing Schafer-Murphy profiles and a modified form of Hill's wife and mother-in-law figure in an attempt to separate the effects of recency and frequency, on perception from those of expectancy.[1] Their *S*s were exposed to different orders of presentation of the unambiguous figures and were so instructed that an expectancy for a given profile was induced on an ambiguous test-trial which followed the series of unambiguous presentations. The 'non-expected' figure was the one most recently seen prior to the exposure of the ambiguous test-figure. These authors found that the responses were in agreement with the most recent perceptual experience on approximately 70–80% of the test-trials. They concluded, therefore, that many perceptual effects, which have been explained in terms of expectancy, should more properly be attributed to the influence of 'recent' memory-traces. Epstein and De Shazo proposed a subthreshold, oscillatory process to explain the means through which recency influences perception and suggested further that this formulation should have general applicability when simple configurations, such as a three-dimensional Necker Cube, were employed.[2]

---

[5] Jones and Marcus, *op. cit.*, 40-44.

[6] See, for example, Fig. 1 in J. E. Stevens, J. D. Mack, and S. S. Stevens, Growth of sensation on seven continua as measured by force of handgrip, *J. exp. Psychol.*, 56, 1958, 246-250.

[1] William Epstein and Irvin Rock, Perceptual set as an artifact of recency, this JOURNAL, 73, 1960, 214-228; Roy Schafer and Gardner Murphy, The role of autism in a visual figure-ground relationship, *J. exp. Psychol.*, 32, 1943, 335-343; E. G. Boring, A new ambiguous figure, this JOURNAL, 42, 1930, 444-445.

Examination of the data obtained from studies employing two-dimensional Necker Cubes create doubt as to the generality of the effect of recency.[3] In these studies, a series of tachistoscopic presentations was first given to determine the perspective preference of each *S* when presented with a 'balanced' Necker Cube. This series consisted of nine presentations of a 'balanced' cube randomly interspersed among 13 right and 13 left 'improved' cubes. Only the data obtained during the preferential trials were examined since they were not contaminated by differential frequency or reinforcement. This situation, then, might be thought of as providing an opportunity to contrast the variables of recency and non-recency. Under these conditions, recency should have even stronger effects than in the studies of Epstein *et al.*[4] The results, however, showed no significant effect of either; they approached a 50-50 division. Of the 215 *S*s taking part in these experiments, 90 gave more responses to recency than to non-recency, whereas 125 *S*s gave more responses to non-recency than to recency. A total of 1935 responses were obtained from these *S*s, and of these 911 were to recency and 1024 were to non-recency. Examination of the data obtained from the first test-trial—which closely approximated the test-condition employed by Epstein and Rock—revealed that 103 responses to recency and 122 responses to non-recency were given.[5]

In addition to these findings, similar results were obtained by Santos with the Schafer-Murphy ambiguous figures.[6] His *S*s gave 141 responses to recency and 147 to non-recency. These data were also uncontaminated by differential frequency or reinforcement.

The evidence presented here relating to the effect of recency suggests that while this may be an important and neglected source of uncontrolled variability in certain perceptual experiments on learning, there is some question as to its generality. Of course, it is possible that the effects obtained by Epstein *et al.* are dependent upon certain procedural factors which are worthy of further investigation if only to provide some information as to when and how this source of variability should be controlled.[7]

---

[2] William Epstein and Del De Shazo, Recency as a function of perceptual oscillation, this JOURNAL, 74, 1961, 215-223; P. A. Adams, The effect of past experience on the perspective reversal of a tridimensional figure, this JOURNAL, 67, 1954, 708-710.

[3] C. M. Solley and J. F. Santos, Perceptual learning with partial verbal reinforcement, *J. Perc. mot. Skills.*, 8, 1958, 183-193.

[4] Epstein and Rock, *op. cit.*, 214-228; Epstein and De Shazo, *op. cit.*, 215-223.

[5] Epstein and Rock, *op. cit.*, 214-228.

[6] J. F. Santos, An investigation of the effects of monetary reward and punishment upon autism in a visual figure-ground relationship. Unpublished master's thesis, Library, Tulane University, 1951.

The Menninger Foundation
Topeka, Kansas

BOBBY J. FARROW
JOHN F. SANTOS

ON VISUAL CHANGES OF REVERSIBLE FIGURES AND AUDITORY CHANGES IN MEANING*

Perspective reversals of static visual figures, like the Necker cube, rotational movements of dynamic visual patterns, like Lissajous' figures;[1] and changes in the meaning of words repeated auditorially,[2] all occur with invariant stimulation. In this sense, at least, they may be considered as analogous phenomena. Do they, however, *(1)* follow similar empirical laws? *(2)* Are they correlated with one another?

To answer these questions, the number of apparent changes per 15-sec. interval during a 3-min. period were obtained from 42 *S*s, who observed the following stimulus materials; (1) a Necker cube with $1\frac{1}{8}$ in. faces; (2) a rotating pin-pattern of the Brown type produced by projecting the shadows of six pins mounted at 60° intervals on the perimeter of a turntable rotating at 15 r.p.m. Dimensions of the shadow-pattern so produced measured $2\frac{5}{16}$ in. wide by $1\frac{5}{8}$ in. high; and (3) and (4) tape recordings of the word 'police' and of the word 'tress' continually repeated.

Question 1. As shown in Fig. 1, rates of apparent change of all four stimulus-materials increased rapidly at first and then tended to level off, confirming in a single group the separate findings of Cohen, Brown, and Warren.[3]

Question 2. The correlations among numbers of changes reported by the *S*s during 3-min. observational period of the cube, the pin-pattern, and the meaning of the words, though all positive, were of such low orders of

---

[7] Epstein and Rock, *op. cit.*, 228; Epstein and De Shazo, *op. cit.*, 223.

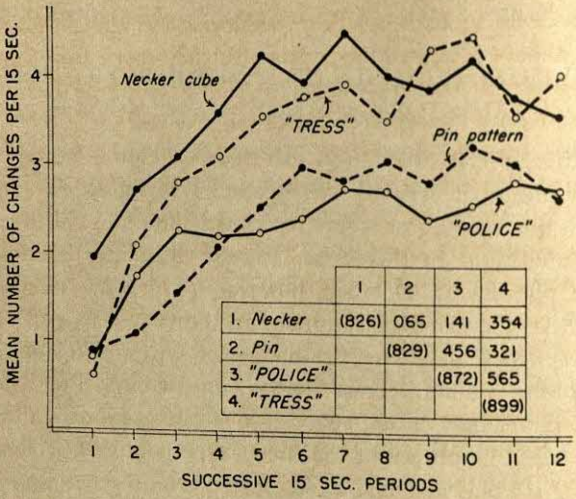
* Supported by National Institutes of Health grants M 900C, M 2109, and H 3582. Thanks are due to Dr. R. M. Warren, who supplied us with copies of some of his tapes.

[1] K. T. Brown, Rate of apparent change in a dynamic ambiguous figure as a function of observation time, this JOURNAL, 68, 1955, 364-368.

[2] R. M. Warren and R. L. Gregory, An auditory analogue of the visual reversible figure, this JOURNAL, 71, 1958, 612-613; R. M. Warren, Illusory changes in repeated words: differences between young adults and the aged, this JOURNAL, 74, 1961, 508.

[3] Leon Cohen, Rate of apparent change of a Necker cube as a function of prior stimulation, this JOURNAL, 72, 1959, 330-332; Brown, *op. cit.*, 364-368; Warren, *op. cit.*, 508.

magnitude that caution must clearly be exercised in thinking of these different types of figure-fluctuations as tapping a unitary process, (*e.g.* 'satiation'). This caveat applies especially (and surprisingly) to considering fluctuations of 'dynamic' visual figures as substantially similar to fluctuations of 'static' ones.



	1	2	3	4
1. Necker	(826)	065	141	354
2. Pin		(829)	456	321
3. "POLICE"			(872)	565
4. "TRESS"				(899)

Fig. 1. Mean Rates of Apparent Change of Necker Cube, Pin-Pattern, 'Police,' and 'Tress' as Functions of Observation Time

$N = 42$ *Ss*. The inset shows the product-moment correlations (decimal points omitted) among numbers of changes of the four stimuli; odd-even reliability coefficients in parentheses: $r_{0.05} = 0.257$, $r_{0.01} = 0.358$ (one-tail).

Duke University

Seymour Axelrod
Larry Thompson

THE FIFTY-FOURTH ANNUAL MEETING OF THE SOUTHERN SOCIETY FOR PHILOSOPHY AND PSYCHOLOGY

The Southern Society for Philosophy and Psychology held its fifty-fourth annual meeting at the Hotel Peabody in Memphis, Tennessee, on April 19, 20, and 21, 1962. Hosts were the Departments of Psychology and Philosophy at the Memphis Naval Air Station, Southwestern at Memphis, Memphis State University, and the University of Tennessee Medical School.

The program included 19 papers and 2 symposia in philosophy, 34 papers and 3 symposia in psychology, and 2 joint sessions. The Junior Award winning papers were presented by W. W. Dawson (U. S. Army Medical

Research Laboratory) on "Examination of the theoretical 'thermal sensation' relationships of Zotterman," and by Donald W. Sherburn (Vanderbilt University) on "Responsibility, punishment, and Whitehead's theory of the self." The Junior Award is a citation for general excellence; it is given annually to the psychologist and the philosopher who are judged to have presented the papers of highest merit from among those papers presented by members who have not yet received the doctoral degree, or who have received it within the preceding five years.

Two joint sessions were pesented. At the first, Rubin Gotesky (Northern Illinois University) served as Chairman, and William M. Hinton (Washington and Lee University), President of the Society, gave the presidential address entitled, "Serendipity in psychology." Iredell Jenkins (University of Alabama) served as Chairman of the second joint session, during which Sigmund Koch (Duke University) presented a paper on "The allures of A-meaning in modern psychology: An inquiry into the rift between psychology and the humanities"; responses to Professor Koch's paper were delivered by E. M. Adams (University of North Carolina) and by Clyde E. Noble (Montana State University), and there followed general discussion from the floor.

Nineteen philosophers and thirteen psychologists were elected to membership during the annual business meeting. The following officers were elected to serve during 1962-1963; Leroy E. Loemker (Emory University), President; Dan R. Kenshalo (Florida State University), President-Elect; Earl A. Alluisi (Lockheed-Georgia Company), Secretary. Leland E. Thune (Vanderbilt University) is in his second year of a three-year term as Treasurer, and William M. Hinton (Washington and Lee University) is the Past-President. Edward A. Bilodeau (Tulane University) and John J. Compton (Vanderbilt University) were elected to the Council as representatives of psychology and philosophy, respectively.

The members accepted the invitation of the University of Miami to hold the Society's fifty-fifth annual meeting at Miami Beach, Florida, on April 11, 12, and 13, 1963.

EARL A. ALLUISI

Lockheed-Georgia Company
Marietta, Georgia

NINTH ANNUAL MEETING OF THE SOUTHWESTERN PSYCHOLOGICAL ASSOCIATION

The Southwestern Psychological Association held its ninth annual meeting at the Hotel Texas, in Fort Worth, Texas on April 5-7, 1961.

Fifty-seven research papers were presented at 14 separate sessions. In addition, there were seven symposia, four workshops, and two invited addresses. The meeting opened with a public symposium on "New directions in mental health," organized and chaired by Ira Iscoe (University of Texas). On Friday afternoon, the presidential address, entitled "What has happened to the other end of the log?—Some observations on the teaching of psychology," was delivered by John P. Anderson (University of Arkansas), and the annual banquet that night had as main speaker J. McV. Hunt (University of Illinois), who spoke on "Intrinsic motivation and experience." At a luncheon meeting on the 7th Robert Voas (Manned Spacecraft Center, NASA) spoke on "The astronaut training program."

A highlight of the meeting was the Annual Competiton for graduate student papers. There were sixteen entries, of whom the following five students were chosen as finalists, reading their papers at a special session on Saturday morning: Thomas W. Hogan (University of Arkansas), Gary L. Holmgren (Texas Christian University), E. Edward Reitman (University of Houston), Joseph A. Sgro (Texas Christian University), and Wayne J. Wilson (Texas Christian University). Winner of the $50.00 first prize was Wayne J. Wilson, whose paper was entitled "The effects of nonresponse—acquisition on latent extinction in a differential reinforcement setting."

At the annual business meeting, it was voted to raise the annual dues to $2.50, with the proviso that a three-year membership may be obtained for $6.00. Also, a resolution was adopted restricting future meetings to locations at which fully integrated convention facilities are available to the membership. The new officers elected for 1962-63 are: Irwin A. Berg (Louisiana State University), President-Elect, and Malcolm D. Arnoult (Texas Christian University), Secretary-Treasurer. Saul B. Sells (Texas Christian University) is current President.

The Association will hold its next annual meeting at the Statler-Hilton Hotel, Dallas, Texas on April 4-6, 1963.

Texas Christian University MALCOLM D. ARNOULT

AN ACKNOWLEDGMENT

The portrait of Professor R. S. Woodworth, which appears as the frontispiece of this number, was sent the JOURNAL by his daughter, Mrs. R. A. Herron. It was taken in the fall of 1956 the day before his 87th birthday. His signature, a facsimile of which appears below the portrait, is a rare sample, as he seldom wrote his full name, being "Woody" to his close friends and "Robt. S." or "R. S." to other acquaintances, was written and sent me at my request during the spring of 1958, when he was in his 89th year.

K.M.D.

Robert Sessions Woodworth: 1869–1962

The death of Robert Sessions Woodworth on July 4, 1962, at the age of ninety-three years, severed one of the very few remaining personal links between the great psychologists of the past and the psychologists of today and those students who will be the psychologists of the future. The significance of Woodworth's long career rests not only on the fact that he knew, or was associated with, or studied with, such psychologists as Wundt, Royce, James, Hall, Külpe, Janet, Münsterberg, Pearson, and Ebbinghaus, with the physiologists, Bowditch, Schäfer, Sherrington and Cannon, and the biologist, Davenport, but also on the fact that he was still teaching at the age of 89 and writing at 91. If one dates the beginning of his interests in psychology from his senior year in college (1891) his active career is seen to cover a span of 71 years. As one of his students put it, "His death isn't the sort of thing one can feel sad about, except as it symbolizes the irrevocable conclusion of a phase of one's life."[1]

Woodworth was born on October 17, 1869, in Belchertown, Massachusetts, of sturdy New England stock. His father, William Walter Woodworth, was graduated from Yale College and Yale Divinity School. He was a minister in the Congregational Church, and except for brief periods of service in Ohio and Iowa, he filled pastorates in New England. His mother, Lydia Sessions Woodworth, was graduated from Mount Holyoke College, and taught there for a time. She was the founder (1856) and first principal of the Lake Erie Female Seminary in Painesville, Ohio, an institution now known as Lake Erie College. An excellent portrait of her hangs in the reception hall of the college. An interesting footnote to his ancestory is that Robert Sessions, the ancestor whose name he bore, was an active participant in the Boston Tea Party. Among his collateral forebears was a Samuel Woodworth (1784-1842), a newspaperman and writer, best known for his song, *The Old Oaken Bucket.*

Robert's father was married three times, and there were children from each marriage. Robert, son of the third wife and born when his father was 55 yr. old, was the eldest of the three of her four children who survived infancy. The household was a large one, with a strong religious atmosphere and prayers and Bible reading twice daily. Discipline was strict, with the father stern and the mother kindly but no less firm. A psychiatrist would be interested in a statement which appears in Woodworth's autobiography. Of his father he said, "As I grew up during his later mellowing years, I became less and less afraid of him."[2] He, himself, questioned the influence of his position in the family upon his personality. Being, as a boy, the youngest among his half-siblings and the eldest among his own living brothers, he never felt sure of the 'true' order. He quotes Adler as indicating that his "style of life" shows that he was an oldest child, yet he himself considered the recent emphasis of psychologists upon "family situation" as overdone, maintaining that, in his own case, influences quite outside his home formed an important part of his environment.

---

[1] When the origins of the quotations and pieces of information are not cited, they come from private files and papers, scattered notes and remembered conversations covering many years.

Ample opportunities offered themselves for a good education. Although in those days scholarships, fellowships and grants were few, and although the family was not particularly well-to-do, funds were forthcoming when needs arose. Woodworth had four years in the excellent high school in Newton, Massachusetts, while living with his sister in that town. His interest during this period varied from farming to music, mathematics, philosophy, astronomy, and the ministry, with music, mathematics, and the ministry preferred. Family influence, though apparently without pressure, gave the ministry a slight advantage. At least it was sufficient to guide him to Amherst as a better place than Harvard for a religiously inclined young man. Amherst was indeed a fortunate choice, for it took him into the classes of Charles Edward Garman, the professor of Philosophy whom afterwards he considered to be one of his most influential teachers. These classes, together with the conferences with their teacher, sharpened his interest in psychology, and showed him the advisability of a thorough grounding in science. He was graduated from Amherst with the A.B. degree in 1891, when his choice of a career narrowed down to the ministry, music, or college teaching.

Advice—probably from Professor Garman—and a need for funds induced him to accept teaching appointments during the next four years, although his achievements in mathematics might well have insured at least a year of graduate work at Harvard. He taught science and mathematics for two years in a high school (1891–92) and mathematics for two more years in Washburn College in Topeka, Kansas (1893–95). He entered Harvard in the autumn of 1895, fairly well committed to a teaching career in philosophy and psychology. There he worked with James,

---

[2] Robert S. Woodworth, in Carl Murchison (ed.), *A History of Psychology in Autobiography*, 2, 1932, 359; also in *Psychological Issues*, 1939, 3.

Royce, Santayana, and Delabarre, and started a life-long and most fruitful friendship with two graduate students, E. L. Thorndike and W. B. Cannon. After he had spent two years at Harvard (A.B., 1896, and M.A., 1897) and had served as assistant in physiology at the Harvard Medical School (1897–98), James McKeen Cattell offered him a fellowship in psychology at Columbia University for the year 1898–99. Woodworth was attracted by Cattell's interest in quantative measurement and particularly by his "mental tests" and their application to the measurement of individual and other differences among people, as exemplified in the so-called Freshman Test Program that had just been introduced at Columbia.

Woodworth was now definitely committed to psychology as a career and at the same time convinced that a thorough grounding in physiology was essential for success. He was awarded the Ph.D. in psychology in 1899, having written a dissertation on "The accuracy of voluntary movement." His special interest was in the relative accuracy of movement when under the control of different sensory mechanisms. He wrote:

> Particular interest attaches to the details of the incoming process, because they condition the details of consciousness. Yet on the other side, the details of the outgoing process are conditioned by consciousness; and consciousness, like other things, should be studied by its effects as well as by its causes. In short, there is no theoretical reason why the outgoing process should not be studied equally with the incoming, by those who are trying to come to an adequate conception of the conscious life of the human organism. We regard this conscious life as built upon the basis of the reflex arc. We have so far studied the afferent portion of this arc. We must advance to the study of the efferent portion. And especially we must endeavor to study the arc as a unit, to trace cause and effect from sense stimulus to muscular response.[3]

The ideas here reflect Woodworth's training in physiology and his dissatisfaction with the current emphasis on sensation in psychology. They foreshadow his concepts as later presented in his *Dynamic Psychology* and other writings.

Following receipt of his doctorate, Woodworth accepted an instructorship in physiology at Columbia and Bellevue Hospital Medical College in New York, and taught there until 1902. The summer of 1900 found him at the University of Edinburgh working in the physiological laboratory of E. H. Schäfer, enjoying the privilege and satisfaction of collaborating with Professor Schäfer on a paper entitled, "Note on the results of circumsection of the motor cortex."[4] He made a short trip to Paris, where he attended the International Congress of Psychology and met many notable Europeans, including Ebbinghaus, Janet, Claparède, and Pearson.

---

[3] The accuracy of voluntary movement. *Psychol. Monogr.* 3, 1899 (No. 13), 2.
[4] With E. A. Schäfer, *op. cit., J. Physiol.*, 26, 1901, xxiii-xxv.

Meanwhile, E. L. Thorndike, whose friendship Woodworth had formed at Harvard, had also gone to Columbia University for a Ph.D. on animal intelligence with Cattell and later an appointment at Teachers College in 1899. The two men collaborated there on an important series of research projects having to do with the doctrine of formal discipline and the concept of transfer of training. There is no way of apportioning the individual contributions of the two in this work, although Thorndike continued to publish actively for years on its educational applications. The first two papers, "The influence of special training on general ability."[5] and "The influence of improvement in one mental function upon the efficiency of other functions."[6] were published jointly.

In these two papers the specificity of the learning process was demonstrated by means of a variety of simple experiments. The "transfer of identical elements" and the "exercise of neural connections" became a part of the literature of psychology and education, and controversy raged for years over the meaning of "identical elements." Two conclusions by the authors show why this came about.

> It is misleading to speak of sense-discrimination, attention, memory, observation, accuracy, quickness, and so forth, since multitudinous, separate, individual functions are referred to by any one of these words. These functions may have little in common. There is no reason to suppose that any general change occurs corresponding to the words, "improvement of the attention," or "of the powers of observation," or "of accuracy. . . ." The mind is, on the contrary, on its dynamic side, a machine for making particular reactions to particular situations. It works in great detail, adapting itself to the special data of which it has had experience. The word *attention*, for example, can properly mean only the sum total of all the particular abilities and inabilities, each of which may have an efficiency largely irrespective of the efficiency of the rest.[7]

An offer of a fellowship at the University of Liverpool in 1902 took Woodworth to England a second time and gave him the cherished opportunity not only to work in the laboratory of Charles Sherrington, where he was appointed Senior Demonstrator of Physiology, but also to be his assistant. He published a paper from that laboratory on "The electric conductivity of mammalian nerve."[8]

Sherrington invited him to continue research in the University laboratories, where he would have been able to develop his experimental psychology and brain physiology together, but an offer of an instructorship from Cattell in the Department of Psychology at Columbia University for the year 1903–1904 proved a stronger attraction. He accepted the post and returned to his familiar surroundings and his friends, this time with a wife, Gabrielle Schjöth, whom he had married in Liverpool. Columbia University became his academic home for the rest of his life, with only temporary absences on sabbatical leaves and for special short assignments elsewhere.

---

[5] With E. L. Thorndike, *op. cit., Psychol. Rev.*, 7, 1900, 140 f. (Abstract).

[6] With Thorndike, *op. cit., ibid.*, 8, 1901, 247-261, 383-395, 558-564.

[7] With Thorndike, *op. cit., ibid.*, 8, 1901, 249 f.; also in *Psychological Issues*, 1939, 337.

[8] *Op. cit., Thompson Yates and Johnston Laboratories Report*, 5, 1903 (Part I), 61-66.

One such assignment came very shortly after his return from England. When the plans for the World's Fair at St. Louis in 1904 included assembling representatives from many different races and made provision for an anthropometric and psychometric evaluation of these samples, it was natural to seek a staff from Columbia University, where Cattell and Franz Boas were already actively involved in measuring individual and group differences. Woodworth, the newly appointed instructor in psychology, was made the director of the project, and a graduate student, Frank G. Bruner, was assigned to him as his assistant. Altogether some 1100 individuals from many primitive races were tested. With a staff of only two examiners, it was decided to give relatively few tests to as many subjects as possible, rather than to make a more intensive examination of fewer cases. The most complete report of the data was made by Bruner, who used the measures of hearing in his Ph.D. dissertation.[9] Woodworth discussed the material in his address as Vice-President of the American Association for the Advancement of Science in Boston in 1909, wherein he foresaw many of the problems and sources of error that arise in dealing with racial differences.[10] The great mass of raw data was kept available in his office, and he would work at it at intervals during the next thirty years. He was, however, too busy to deal with the tremendous number of correlations needed, for there were then no such computing devices as are available today. Further, he recognized that the great developments in test materials and techniques in the intervening years had probably made the data obsolete. Nevertheless he had demonstrated to psychologists and anthropologists alike the feasibility and value of systematic measurement of racial and group differences. His continuing interest in racial differences was evident in many ways, particularly in his long editorship of the sections on "Race and Individual Psychology" for *The Psychological Bulletin*.

The work of the new instructorship at Columbia proved difficult and taxing. Facilities and space were meager, and the budget of the Department was low, as indeed might be expected in a department within the

---

[9] F. G. Bruner, The hearing of primitive peoples, *Arch. Psychol.*, 2, 1908 (No. 11), 1-113.

[10] Racial differences in mental traits, *Science*, 31, 1910, 171-186.

Faculty of Philosophy where paper and ink are the traditional tools. Moreover, Woodworth's chief, Cattell, had an understanding with the University that he himself was to be at the University only two days each week. That left the housekeeping and administrative duties, as well as much of the teaching and supervision of students, to Woodworth. His performance must, however, have been acceptable for he was promoted to an adjunct professorship in 1905, and to a professorship in 1909.

His first formal appearance before a University assemblage was in 1908, whene he delivered a lecture on "Psychology" in a series known as Science, Philosophy, and Art. He started his lecture thus:

> In attempting to give a view of the present state of his science . . . the psychologist encounters a peculiar difficulty in that the camp of his fellows is divided. Two standards are displayed; two parties are in the field; and their relations are not always as friendly as might be expected of soldiers in a common cause. Such divisions are indeed not unknown in other sciences. The descriptive and dynamic phases of a subject, the study of structure and the study of function, frequently divide the workers between them, and sometimes those who hold to the one despise the other. The line of cleavage is sharper in psychology than elsewhere: those who define psychology as the science of consciousness . . . maintain that any study which does not contribute to the description of conscious life lies beyond the pale of the science; whereas to those who define it as the science of those functions of the organism which are roughly designated as *mental*, the examination of the consciousness attending these functions, though valuable and suggestive, do not lie at the root of the matter.[11]

There followed a discussion of objective and subjective psychology of introspective and behavioristic psychology. Woodworth took an intermediate stand, advocating the use of whatever methods and materials would give the best explanation. It is interesting to compare this lecture of 1908 with one twenty-three years later given before an equivalent audience and on the same topic.[12] Each is a thorough exposition of psychology as it then was. One of his opening remarks was:

> It was a surprise to me just the other day, on reading over this lecture of 1908, to find the cleavage of the introspective and objective wings of psychology so much to the fore, several years before the emergence of behaviorism.

It should be no surprise that he could have foreseen so clearly the course that psychology was to take in the following quarter century.

The American Psychological Association in 1906 appointed a committee to serve as a general control committee on the subject of measurements. This committee in turn appointed a sub-committee on Association Tests, whose members were Woodworth and F. L. Wells. Its report became a well-known source of test material, and is still referred to today.[13] One could find in every psychological laboratory The Color-Naming Test, The Form-Naming Test, The Logical Relations Test, and many others, together with a test to measure the understanding of instructions. This last one is of particular interest as it set a pattern for innumerable variations in modern tests of intelligence and special abilities. The students in the Columbia laboratories were kept busy for many months taking these tests, making critical comments upon them, and frequently administering them in the process of standardization.

---

[11] *Psychology,* Columbia University Press, 1908, 5.

[12] Psychology, in D. R. Fox (ed.), *A Quarter Century of Learning, 1904-1929,* 1931, 129-146.

Overlapping this research on Association Tests was Woodworth's part in the revision of G. T. Ladd's *Physiological Psychology.* This book was first published in 1887, three years before James' *Principles,* and had gone through some 10 reprintings without revision. Woodworth was invited to revise the sections dealing with the nervous system, an undertaking which turned out to be a stupendous job, involving the complete rewriting and expansion of a large proportion of the book. In fact Ladd, to keep the whole within acceptable dimensions, reduced his section on psychology to two chapters, with the following comment:

> The two chapters to which this part had been reduced . . . simply maintain the contention that the study of ultimate psychological problems, from the physiological and experimental points of view, finally leaves them in essentially the same condition as that in which it finds them, and this position is that of a naive and so-called "common-sense" dualism, as distinguished from the metaphysical theories of parallelism, materialism, or subjective idealism. . . . During the progress in preparation of this edition, the author of the first edition found that his obligations to Professor R. S. Woodworth could be justly discharged only by placing his name upon the title-page as its joint author. And this act of justice was cheerfully rendered.[14]

Thus was born Professor Woodworth's first major publication. It was known generally as "*Woodworth's* Physiological Psychology," and in many places is still in use.

One of his most appreciated honors, according to Woodworth himself, came to him in 1914 with his election as President of the American Psychological Association. He always was, before that date and up to the time of his death, a loyal and enthusiastic member, attending its annual meetings whenever possible. The office of President carried great prestige, and the rivalry for election was keen. Some inkling of this contest was evident, when, after his election, he remarked, with a twinkle in his eye,

---

[13] With F. L. Wells, Association tests, *Psychol. Monogr.,* 13, 1911 (No. 57), 1-85.

[14] G. T. Ladd and R. S. Woodworth, *Elements of Physiological Psychology,* rev. ed., 1911, 7 f.

"Well, anyway, I was elected at least a year before Watson!" His presidential address[15] gave him the opportunity to develop further his conception of imageless thought which he had discussed in earlier papers[16] and which had been in his mind since he wrote his dissertation on "The accuracy of voluntary movement" in 1899. Woodworth was not a crusader for causes, and this argument is about the closest he came to championing a cause. His contemplative attitude led him to see all sides of a problem and usually to come up with a compromise. Earlier he had said:

> I was thus led to make the statement that imagery, when present, was but the clothing of the thought, and that a naked thought was fully capable of doing the work.[17]

In his revision he made the following pertinent comment and seemingly disposed of the matter:

> In imageless thought, then, the imagelessness has nothing particular to do with the thinking process; and we are permitted to drop, with some relief, the elevated tone that has sometimes seemed appropriate to the topic. Thought is imageless because its data are recalled in an imageless form, and not because it does not thrive in a sensory atmosphere.[18]

In the academic year 1916–17, Woodworth was invited to give a series of lectures at the American Museum of Natural History under the auspices of the Museum and with the coöperation of Columbia University, a series entitled *Dynamic Psychology*. These lectures were prepared against great pressures of teaching and of the direction of student research. On some occasions the last page would be pulled from the typewriter followed by a dash to the lecture hall, or an interruption would prevent the last part of the lecture from being written down at all. Nevertheless, the project served a most useful purpose, for it forced Woodworth to organize ideas that had been in his mind for many years. He was strongly negative to those "who try to prescribe what one must do or the results one must find," and his *Dynamic Psychology* was his reaction against Titchener, Watson, and McDougall. In it he "sought for a position that should be independent and yet have room for all genuine psychological efforts." The following enlightening paragraph is found in his autobiography:

> My bogey men—the men who most irritated me, and from whose domination I was most anxious to keep free—were those who assumed to prescribe in advance what type of results a psychologist must find, and within what limits he must

---

[15] A revision of imageless thought, *Psychol. Rev.*, 32, 1915, 1-27; also in *Psychological Issues*, 103-127.

[16] Imageless thought, *J. Philos.*, 3, 1906, 701-707; also in *Psychological Issues*, 72-79.

[17] Imageless thought, *J. Philos.*, 3, 1906, 702; also *Psychological Issues*, 73.

[18] A revision of imageless thought, *Psychol. Rev.*, 32, 1915, 10; also *Psychological Issues*, 111.

> remain. Münsterberg was such a one, with his assertion that a scientific psychology could never envisage real life. Titchener was such a one, in insisting that all the general findings of psychology must consist of sensations. Watson was such a one, when he announced that introspection must not be employed, and that only motor (and glandular) activities must be discovered. I always rebelled at any such epistemological table of commandments.[19]

The eight lectures at the American Museum of Natural History were published in book form by the Columbia University Press in 1918 under the title *Dynamic Psychology.*[20] It went through a number of printings without revision and had a moderate but steady sale for years. The Nicholas Murray Butler Medal (in silver) was awarded to Woodworth for his distinctive contribution to philosophy, psychology and education, following the publication of the book.

Dynamics of Behavior, published in 1958, is in a sense a successor to the earlier book, in that it preserves the same general point of view, but it is in no sense merely a revision either in content, size, or arrangement.[21] It embodies the lectures given in a course at the University under that title.

When the United States entered the First World War, Woodworth, aged forty-seven years, was persuaded not to enter the Psychological Service, as large numbers of his colleagues were doing, but to do his duty as a civilian. At once he was commissioned by the American Psychological Association to attempt the preparation of a test for emotional stability, a test badly needed in the evaluation of recruits. With the help of his colleagues, he collected early specific symptoms of neurotic tendencies by examining hundreds of printed case histories, and talking with psychiatrists, neurologists and clinical psychologists. These symptoms were cast into the form of simple questions that could be answered "Yes" or "No." All "symptoms" which showed up with high frequency in the large number of supposedly normal persons who answered the questions were eliminated as probably having no diagnostic value. The remainder, after repeated weeding out and simple statistical treatment, became the Woodworth Personal Data Sheet. Too late to be thoroughly tested on recruits for its diagnostic value, it became a device widely used by civilians. It was not copyrighted and, although it generally retained the name, *Personal Data Sheet,* its origin dimmed with passing years. It, or parts of it, are still in use in one or another modified form, and under a variety of names. It has undoubtedly become a most important, although entirely unanticipated, contribution to the field of personality measurement.

---

[19] Autobiography, *A History of Psychology in Autobiography,* 2, 1932, 376; also *Psychological Issues,* 1939, 21.

[20] *Dynamic Psychology,* 1918, 1-210.

[21] *Dynamics of Behavior,* 1958, 1-403.

An inventory of the courses taught by Professor Woodworth over the years leads to the conclusion that invariably "he did what needed to be done." Although physiology and experimental psychology were his basic courses, at one time or another he taught an introductory course, abnormal psychology, social psychology, tests and statistics, problems and methods, theory, history, and applications of psychology, and a survey of contemporary psychology, as well as seminars for advanced students on many and varied topics. From a number of these courses there came important books. In addition to *Dynamic Psychology* and *Physiological Psychology,* there were an introductory textbook called simply, *Psychology,*[22] which went through four revisions (with a fifth revision in collaboration with D. G. Marquis) and which has been generally credited with being the most widely used elementary text; the big *Experimental Psychology* first published in 1938,[23] and revised in 1954, under joint authorship with Harold Schlosberg; and *Contemporary Schools of Psychology,*[24] first published in 1931, largely rewritten in 1948, and in the process of being rewritten again shortly before his death.[25]

Woodworth never spared himself in his service to his students. In one of his frequent self-appraisals, he said (in 1928) that he was inclined to leave his students pretty free, and to be sought rather than to seek them out for consultation and direction. He also said that he thought he probably gave too much time to each student and did not discriminate well enough between "worthwhile students" and those who were not.

One of his too little appreciated services to his students, and indeed to psychology in general, was his editorship of the *Archives of Psychology.* This series of monographs he edited from 1906 until 1948, when it was presented to The American Psychological Association. In the earlier years he took full responsibility for the editorial work and the financial accounts with no clerical assistance whatsoever. Columbia University required full publication of all doctoral dissertations and the deposit of 75 copies of each with the University Library. The monograph series provided a ready medium and frequently the only medium for satisfying the requirement at a cost that the student could afford. A copy of the following letter is in Woodworth's personal file:

---

[22] *Psychology,* 1921, 1-580; rev. ed., 1929, 1-590; 3rd ed., 1934, 1-546; 4th ed., 1940, 1-639; with D. G. Marquis, 5th ed., 1947, 1-677.

[23] *Experimental Psychology,* 1938, 1-889; with Harold Schlosberg, rev. ed., 1954, 1-948.

[24] *Contemporary Schools of Psychology,* 1931, 1-232; rev. ed., 1948, 1-279.

[25] The revision will be completed by Dr. Mary Rose Sheehan.

August 18, 1939

To Whom It may Concern

> I do not regard the *Archives of Psychology* as my personal property. I regard its funds as in effect a trust fund to be devoted for publishing Psychological Research. Its present balance has been built up gradually over many years. None of my personal funds are included in the *Archives* balance.
>
> (Signed) Robert S. Woodworth

From his student days, and like most psychologists with training in philosophy, Woodworth had a deep interest in the relationship between the psychical and the physical. He often recalled his conversations concerning the mind-body relationship with William James, who was an "interactionist." Neither this theory nor any one of several others satisfied Woodworth. He finally worked out an interpretation which seemed satisfactory to him personally although he said he doubted that it would satisfy many others. It was a form of "parallelism" which he called "the parallelism of sciences," really a parallelism of different scientific descriptions of the same processes. For example:

> If a subject under suitable acoustical stimulation reports hearing a tone, and a physiologist inspecting his brain reports certain detailed processes as occurring, there is no longer any doubt in my mind that these detailed processes are parts of the identical total process which the subject himself reports.[26]

Woodworth had a wise insight into his own personality, nor was he averse to rating his shortcomings. He was not a good administrator and had a definite sense of inferiority on "the executive and achievement side." His tendency toward "mature rumination," together with his need for extended contemplation of a problem, led him to delay decision and action in administrative matters. Indeed he was the antithesis of a good administrator, who has somewhere been defined as "one who can make decisions upon insufficient evidence." It was a great satisfaction to him when, in 1925, he was relieved of a large part of the burden of administration and could at last use his time more nearly as he wished.

He was not, nor did he want to be, a leader. In 1957 he wrote that he had always submitted readily to being led and directed by those superior in position or by those he recognized as good leaders. He disliked assuming a dominating role but preferred at most a relationship of equality between himself and his co-worker. He wrote a personal note in 1928, in which he said:

> I rate my achievement very low, believing that I am one of the sort whose name will soon be forgotten, though of service during his lifetime. Such things as I have achieved have probably been in the way of contributing to discussions and assisting

---

[26] Autobiography, *A History of Psychology in Autobiography,* 379; *Psychological Issues,* 24.

> to clarify issues. I seldom have carried the ball, but have often taken part in the interference.

At another time he said:

> I have made no discoveries of any importance. I have invented certain useful experimental techniques, directed many research students, and believe I have made some contributions to clarity of psychological findings and conclusions by criticism and synthesis.

The following few words, written in 1935, show his lasting faith in what science can do when it is properly applied to the world's pressing problems:

> My creed in regard to current public and economic questions amounts to little more than this: that our present and traditional arrangements, far from being divinely ordered, are simply the best muddling humanity has managed to achieve thus far, and that fair-minded investigation can be trusted to show the way to better things.

The many honors that came to him, particularly in his later years, give ample evidence that he grossly underrated his achievements and his standing among his fellows in the scientific world. What are honors to any given person and what their significance is to the recipient can only be inferred by others and frequently with a large margin of error. Two that were paid staff jobs were nevertheless signal honors to Woodworth. The first was his selection as Chairman of the Division of Anthropology and Psychology of the National Research Council in 1924-25, where he met with the outstanding men of the physical and biological sciences along with those in the fields of psychology and anthropology. In addition to the administrative duties of the office, duties which were not onerous, the chairman had the opportunity to develop those projects in which he was particularly interested. Woodworth worked on two of these: The committee for Research in Child Development, which some ten years later grew into the important and active Society for Research in Child Development, with its own publications, *Child Development, Abstracts of the Society for Research in Child Development* and *Monographs of the Society for Research in Child Development*. The other was the Committee on Psychology of the Highway, which is still an active and important factor in improving highway safety.

The second honor was his selection as President of the Social Science Research Council in 1931–32, after having served as member of its Board of Directors, its Treasurer, and as a member of a number of its important committees. He felt it an honor, not only to himself, but to the field of psychology to be placed in a working relationship with the leaders in economics, history, political science, sociology, and statistics. As an outgrowth of his membership on its Social Adjustment Committee, he prepared a monograph on *Heredity and Environment: A Critical Survey of*

Recently Published Material on Twins and Foster Children.[27] This monograph has more than fulfilled the expectation that it would "be helpful not only to those engaged in research on twins and foster children but also to all interested in the role of the genetic and environmental factor in human behavior."

When Cattell established the Psychological Corporation in 1921 "for advancement of psychology and the promotion of the useful applications of psychology," Woodworth became one of the original members of the Board of Directors and was elected president of the Corporation in 1929. He attended the regular Board meetings until 1960, when he was appointed to the honorary position of Director Emeritus. He served on many of the Corporation's committees, which profited from his wise and balanced counsel.

His colleagues in Columbia University on the occasion of his seventieth birthday (1939) presented him with a volume containing a selection of his papers, with the title of *Psychological Issues.*[28] On the occasion of his ninetieth birthday a group of his students and colleagues published *Current Psychological Issues* under the editorship of two of his former students, G. S. Seward and J. P. Seward, "as a tribute to this outstanding pioneer who has opened up so many new pathways in the science to which he has devoted his long, productive life."[29]

He was a member of Phi Beta Kappa and Sigma Xi, a fellow of the American Academy of Arts and Sciences, a member of the American Philosophical Society and of the National Academy of Sciences.

The honorary degree of Sc.D. was conferred upon him by Columbia University in 1929, by the University of North Carolina in 1946, by the University of Pennsylvania in 1946, and by Amherst College in 1951. He was given an LL.D. by Lake Erie College in 1947.

It seems fitting to close this brief appreciation of Professor Woodworth by one who was intimately associated with him as student, colleague, and friend for fifty-three years, by citing the crowning award to him on September 2, 1956, of the first Gold Medal of the American Psychological Foundation, "for distinguished and continuous service to scholarship and research in Psychology and for contributions to the growth of Psychology through the medium of scientific publication."

Columbia University

Albert T. Poffenberger

---

[27] Heredity and environment: A critical survey of recently published material on twins and foster children, *Soc. Sci. Res. Coun. Bull.*, 1941 (No. 47), 1-96.

[28] *Psychological Issues,* published by a Committee of his Colleagues, Columbia University, 1939, 1-410, which contains his bibliography to that date.

[29] G. S. Seward and J. P. Seward (eds.), *Current Psychological Issues: Essays in Honor of Robert S. Woodworth,* 1958, 1-360.

A BIBLIOGRAPHY OF ARTICLES AND BOOKS BY ROBERT SESSIONS WOODWORTH: A CONTINUATION, 1938-1959

By ENRICA H. B. TUNNELL, Columbia University

A bibliography of 198 articles and books by Robert Sessions Woodworth from 1897 through early 1938 appears in *Psychological Issues: Selected Papers of Robert S. Woodworth* (New York, Columbia University Press, 1939). Here follows a continuation from 1938 to 1959 containing 48 additional items. Three publications omitted from the earlier bibliography in *Psychological Issues* are given initially under their respective years. For conformity, the style of the first bibliography is followed.

1907

El Movimiento. (A Spanish translation by Domingo Vaca of *Le Mouvement,* 1903.) Madrid, Daniel Jorro, 448.

1926

[*Behavior Psychology*]. (A Japanese translation by N. Asahi of *Psychology,* 1921.) Dairen, Manchuria, 1926, 446.

1933

Christine Ladd-Franklin, *Dictionary of American Biography,* 10, 1933, 528-530.

1938

Experimental Psychology. New York, Holt, 1938, 889.

Psykologia. (A Finnish translation by U. Harva and E. S. Saarimaa of *Psychology,* 3rd ed.) Helsinki, Werner Söderström, Osakeyhtiö, 1938, 558.

1939

Fundamental and applied research, *Mental Health,* Publication No. 9 of the American Association for the Advancement of Science, 1939, 59-61.

Individual and group behavior, *Amer. J. Sociol.,* 44, 1939, 823-828.

Psychological Issues. New York, Columbia University Press, 1939, 421.

1940

Charles August Strong: 1862-1940, *Amer. J. Psychol.,* 53, 1940, 302.

(With M. G. Smith). *Class Demonstrations in Psychology.* New York, Holt, 1940, 80.

Livingston Farrand: 1867-1939, *Amer. J. Psychol.,* 53, 1940, 302.

(With M. G. Smith). *Notes for Instructors, to Accompany Class Demonstrations in Psychology.* New York, Holt, 1940, 12.

Psychology. 4th ed. New York, Holt, 1940, 639.

Recent results on heredity and environment, *Trans. N.Y. Acad. Sci.,* 3, 1940, 30-35.

A review of Thorndike's *Human Nature and the Social Order, Science,* 92, 1940, 36-38.

A Revised Standardized Examination on Woodworth's Psychology. 3rd ed. Forms A and B. New York, Holt, 1940, 56.

1941

Heredity and environment: A critical survey of recently published material on twins and foster children, *Soc. Sci. Res. Coun. Bull.,* 1941 (No. 47), 1-96.

A review of Cattell's *General Psychology, Science,* 94, 1941, 116-117.

Successes and failures of experimental psychology, *Science,* 94, 1941, 265-270.

1942

The Columbia University Psychological Laboratory: A Fifty-Year Retrospect. New York, Author, 1942, 23.

Introduction, *Studies in Personality: A Contributions in Honor of Lewis M. Terman*, ed. by Quinn McNemar and M. A. Merrill, 1942, 3-12.

Raymond Dodge: 1871-1942, *Psychol. Rev.*, 49, 1942, 395-402.

A review of Boring's *Sensation and Perception in the History of Experimental Psychology*, *Science*, 96, 1942, 64-65.

A review of Miller's *Unconsciousness*, *Science*, 96, 1942, 114.

1943

The adolescence of American psychology, *Psychol. Rev.*, 50, 1943, 10-32.

Bugünün Psikologi Cereyanlari. (A Turkish translation by Mazafer Sherif [Başoğlu] of *Contemporary Schools of Psychology*, 1931.) Ankara, Ankara University, 1943, 210.

1944

(With M. R. Sheehan). *First Course in Psychology.* (Trade edition: *Practical Psychology of Everyday Life.*) New York, Holt, 1944, 445.

(With others). James McKeen Cattell: In Memoriam, *Science*, 99, 1944, 151-165.

James McKeen Cattell: 1860-1944, *Psychol. Rev.*, 51, 1944, 201-209.

[*Psychology*]. (An Arabic translation by Abdul Hamid Kadhim of *Psychology*, 4th ed.) Baghdad, Iraq, 1944, 788.

1945

Edmund Burke Delabarre, *Science*, 102, 1945, 369.

1946

Reinforcement of perception, *Amer. Psychologist*, 1, 1946, 260-261 (abstract).

1947

Reinforcement of perception, *Amer. J. Psychol.*, 60, 1947, 119-124.

(With D. G. Marquis). *Psychology.* 5th ed. New York, Holt, 1947, 677.

1948

Contemporary Schools of Psychology. Rev. ed. New York, Ronald Press, 1948, 279. 1948, 279.

Margaret Floy Washburn: 1871-1939, *Nat. Acad. Sci. Biogr. Mem.*, 25, 1948, 275-295.

1949

Psychologie experimentale. (A French translation by André Ombredane and Irène Lézine of *Experimental Psychology*, 1938.) Paris, Presses Universitaires de France, 1949, 2 vols., 1183.

(With M. R. Sheehan). [*Practical Psychology of Everyday Life*]. (A Hebrew translation by M. Kokhba of *First Course in Psychology*, 1944.) Tel-Aviv, Masadah, 1949, 262.

A review of Hilgard's *Theories of Learning*, *J. abnorm, soc. Psychol.*, 44, 1949, 124-129.

1950

Edward Lee Thorndike: 1874-1949, *Science*, 111, 1950, 250-251.

1951

Psychologie. (A Dutch translation by P. P. J. Van Caspel of *Psychology*, 5th ed.) Antwerp, Standaard-Bockhandel; Utrecht, Spectrum, 1951, 693.

(With M. R. Sheehan). *First Course in Psychology.* Rev. ed. New York, Holt, 1951, 445.

1952

Edward Lee Thorndike: 1874-1949, *Nat. Acad. Sci. Biogr. Mem.*, 27, 1952, 209-237.

1953

[*Psychology*]. (A Hebrew translation by M. Elan-Rosenhach of *Psychology,* 4th ed.) Tel-Aviv, Masadah, 1953, 590.

Psychkologi. (A Swedish translation by Elsa Janson and Mils-Gustac Hildeman of *Psychology,* 5th ed.) Stockholm, Almqvist & Wiksell, 1953, 960.

1954

(With Harold Schlosberg). *Experimental Psychology.* Rev. ed. New York, Holt, 1954, 948.

[*Contemporary Schools of Psychology*]. (A Hebrew translation by B. Harel of *Contemporary Schools of Psychology.* Rev. ed.) Tel-Aviv, Hebrew Teachers Union, 1954, 216.

1958

Psicologia. (A Portuguese translation by Lavinia Costa Raymond of *Psychology,* 5th ed.) Sao Paulo, Campanhia Editors, National, 1958.

Dynamics of Behavior. New York, Holt, 1958, 403.

1959

John Broadus Watson: 1878-1958, *Amer. J. Psychol.,* 72, 1959, 301-310.

Josiah Royce: 1855-1916, *Nat. Acad. Sci. Biogr. Mem.,* 33, 1959, 381-396.

BOOK REVIEWS

Edited by T. A. Ryan, Cornell University

Handbook of Physiology. A Critical Comprehensive Presentation of Physiological Knowledge and Concepts. Section 1: Neurophysiology. Vols. I, II, III. Edited by John Field (Section editor, H. W. Magoun, Executive editor, Victor E. Hall). Baltimore, Md., Williams and Wilkins Co., 1959 (Vol. I), 1960 (Vol. II, III). Pp. 2013. Vol. 1, $22.00; Vol. II and III, $20.00 each.

The psychiologist will find these three volumes indispensable as source material for research and especially for graduate teaching. There are 81 chapters by 93 authors including 11 psychologists. In fact, a selection of chapters will provide an ideal textbook of physiological psychology for the advanced student.

Following Sherrington's classical plan the text portrays present knowledge of the 'way in,' the 'way out,' and the central integrative mechanisms of the nervous system. But this rigid didactic schema has been drastically modified in the past 25 years. Contemporary research has added information about feed-back mechanisms which delicately regulate in-going patterns of nerve impulses as well as the precise adjustments of effector action. R. B. Livingston says in Chap. XXXI:

> The nervous system appears to be made up less of independent linear pathways than of mutually interdependent loop circuits which stick together the various parts of the brain into a functional whole. Along ascending as well as descending projections, the brain-stem reticular formation and the cerebral and cerebellar systems linked closely with it seem to modulate impulse traffic in a continuous action that modifies the composition of perceptive as well as projective neural patterns. The losses and distortions of signals brought about by this mechanism favor the conclusion that some teleological mechanism is at work; this appears to be designed to diminish the involvement of higher centers with signals that have little immediate significance for the animal. . . . A further examination of these mechanisms will help us to understand many problems of absorbing interest in neurology, psychiatry, psychology, and philosophy (p. 759).

This work is the product of an editorial *tour de force*. Each chapter is lucid and beautifully illustrated. Moreover, integration is furthered by copious cross references from chapter to chapter and by meticulous bibliographical references for each of the topics discussed.

Of particular interest to psychologists are the opening and closing chapters and the nine introductory chapters written by such eminent investigators as E. D. Adrian and Wilder Penfield. The opening chapter: "The historical development of neurophysiology," by Mary A. B. Brazier (40 illustrations) and the closing chapter: "Neurophysiology: An integration," by R. W. Gerard would be most useful to graduate students of psychology if issued as an offprint including the nine brief introductory chapters. This concentrated bit of knowledge could be included in a booklet of 158 pages.

The "critical comprehensive presentation of physiological knowledge" here presented is but one installment of a continued story. It is the intention of the American Physiological Society "to cover the physiological sciences in their entirety once in about 10 yr. and to repeat the process periodically thereafter." Previous handbooks of physiology were mostly written in German. The only compendium in English was E. A. Schäfer's *Textbook of Physiology,* 1898-1900. Since it has been estimated that our scientific knowledge doubles every 15 yr. it is interesting to compare the neurophysiological portion of Schäfer's textbook with the present three volumes of 2000 pages. Schäfer's book contains 906 pages (15 chapters) on the nervous system by nine authors including but one psychologist (W. H. R. Rivers writing on vision). Schäfer limited all biochemical information to Vol. I, "Chemical Physiology." By contrast, in the present handbook relevant biophysical and biochemical knowledge is closely integrated with neural function in all three volumes. However, 9 chapters in Vol. III are exclusively devoted to chemical aspects of neural function (Chaps. 70-78).

Looking back upon 45 yr. of laboratory investigation, first in psychology then in physiology, the reviewer wishes to make a few critical comments on changes in the 'climate' of experimental research in these fields as he has experienced them. Beginning with the present state of neurophysiological research as reviewed in these excellent, authoritative chapters, let us turn our attention to three of them of particular interest to psychologists; namely, the neural basis of learning (Chap. 61), emotional behavior (Chap. 63), and psychosomatics (Chap. 69). These topics have been chosen for scrutiny because the reviewer's own laboratory perspective has centered on this area.

Galambos and Morgan (Chap. 61) in their concluding remarks say:

> The major obstacle (to knowledge about the neural basis of learning) has undoubtedly been the inability to break through the barrier of the calvarium in order to expose the brain of normal unanesthetized animals to direct experimental investigation. Within the last few years this barrier has been effectively overcome. Wires can now be placed in practically any desired brain structure, and consequently the study of electrical responses of the brain is no longer limited to the pale and distorted picture provided by scalp electrodes. . . . On the behavioral side the development of new techniques of measurement (*e.g.* the operant methods of Skinner) and attention being paid to types of behavior not previously examined (*e.g.* imprinting) have appreciably enlarged both the precision and scope of the analysis. . . . It is no longer fashionable to conceive of the 'temporary connections' as occurring exclusively in the cortex. At least the reticular substance and the limbic system can in addition be presumed to play important roles, and thus two entirely new parameters are provided along which experimental attacks can be made.
>
> The present era in neurophpsiology is therefore unlike any other with respect to the neural basis of learning. As recently as 5 years ago an observer not ordinarily given to pessimism might well have concluded from the available evidence that our search for the answer was permanently doomed. But today, when obviously relevant measurement of many types are being made in the brain substance of unanesthetized animals, an avalanche of entirely new information is about to descend upon us and such a conclusion is hardly justified.

In Chap. 63, Brady's account of emotional behavior is placed within this same laboratory framework. He and contemporary investigators are employing electrical stimulation of the brain (including the animal's self-stimulation) and the graphic registration of motor responses, often of many individuals simultaneously. 'Psychosomatics' has now been officially recognized by the editors as a branch of neurophysiology and is reviewed by MacLean in Chap. 69.

Now for a brief excursion to sample the scientific 'climate' in psychology and physiology of the early twenties, as this reviewer experienced it. In Professor

Titchener's laboratory in 1918 graduate students were required to take the course in physiology with the first year medical students under Professor Sutherland Simpson, a pupil of E. A. Schäfer. We were introduced by Titchener to the practice of rigidly controlled introspection and in physiology to the experimental tradition of Sherrington and the beginnings of endocrinology. (E. A. Schäfer and Sutherland Simpson were pioneers in this new field).

Since each graduate student served as subject in all current problems, he acquired skill in introspection under experimental control in a variety of human experiences. Titchener's aim was to decipher by trained introspection the details of mental structure and to correlate these anatomical details of mental life with the accompanying physiological processes. This arduous and tedious task led to the publication of an impressive body of introspective data. For example, Titchener's eminent pupils, Karl M. Dallenbach and E. G. Boring contributed to the analysis of the experiences of being awakened in the night to guess the time, of tactile stimulation (including the return of sensation following cutaneous nerve section), and of visceral processes such as hunger, nausea, defecation, and micturition.

But where were the physiological correlates to be found in the central integrative mechanisms of the nervous system? They are only recently being disclosed. Brazier (Chap. 1, p. 56) says:

> It is the great advance in electrical stimulation and recording that marks this era of investigation from its predecessors, although it is only through knowledge from all sources that progress can be achieved in an understanding of the brain. . . . The search began for the physiological mechanisms of mental processes, of consciousness, of memory . . . all terms and concepts that had belonged to another domain of thought.

This reviewer first became aware of Pavlov's contribution to psychology in teaching medical students the physiology of digestion in 1920. Through the detailed examination of psychical secretion Pavlov came to devise his classical method of the conditioned reflex. Through this pedagogical approach (teaching about digestion), the reviewer adopted the conditioning method in his own research and found himself back in psychology in 1923.

Contemporary neurophysiologists and psychologists are disposed to avoid Pavlov's laborious time-consuming and restricted method of the conditioned reflex. (This, of course, is not true in the communist countries). Brazier (Chap. 1, p. 56) says:

> Pavlov's ideas of the reflex became more diffuse as he grew older. Experiments to test the modes of behavior of animals to conditioning stimuli were less difficult to design than ones to test the hypothesis advanced to explain them. Temporary neuronal connections in the cortex proved easier to postulate than to prove . . . and it is only in recent times that the electrophysiologist's tools have been applied to the problem.

Titchener's difficulty all over again!

Let us now return to the present. The 'blind spot' in the contemporary neurophysiology of learning and emotion is, as I believe, mainly due to the emphasis on short-term behavior, as John Lilly calls it. This emphasis results from the present-day formidable armamentarism of exquisitely precise surgical, electrical, and chemical techniques together with a confusing array of animal training methods. This embarrassment of riches inclines the enthusiastic young investigator to brief spurts of focused zeal in his observations of behavior. Moreover, the automatic registration of the performance of numerous animals at the same time involves the *anonymity of* *the experimental animal.* This anonymity is impossible with Pavlov's tedious method of the conditioned reflex in which, after months or even years of training the animal emerges as an *individual* and in so doing develops a 'patient-physician' type of relationship with the experimenter. Pavlov's observations of experimental neuroses in his dogs, subjected to prolonged and arduous conditioning, led directly to present research in animal and man in the new field of psychosomatics.

In characterizing psychosomatics and in closing this review, MacLean (Chap. 69, p. 1729) says:

> Neurophysiology, with its access to animal experimentation, is peculiarly suited to making a contribution to psychosomatics. In order to indicate the nature of the problems in which the aid of neurophysiology is primarily required, it will be necessary again to appeal to introspective material. On the basis of such material it is recognized that emotion is the only form of psychological information which short of physical exercise, is associated with extensive internal changes of the body. In view of this, it is a matter of first importance to ascertain whether or not one can localize mechanisms in the central nervous system that are particularly concerned with the experience of emotion and its elaboration into behavior. . . . *Another major problem in line with this approach is to discover whether or not emotion and its underlying central and peripheral mechanisms can initiate internal changes that are either sufficiently intense or enduring to result in lesions* (Reviewer's italics).

Cornell University HOWARD S. LIDDELL

The Sources of Value. By STEPHEN C. PEPPER. Berkeley, California, University of California Press, 1958. Pp. xiv, 732. $10.00.

No one whose interest in any way touches the field of human behavior can afford to be ignorant of this book.[1] Though the author modestly calls it a "sketch," he presents a comprehensive hypothesis about the facts bearing on the main problem in value theory from Plato to the present: "how to make well-grounded decisions in human affairs." Granted this aim, the book is a resounding success. It is also a fine example of expository writing.

The pivotal concept in Pepper's value theory is that of "selective system." Anything chosen or rejected by a selective system is valued positively if it is chosen, or valued negatively if it is rejected. There are as many different classs of value as there are selective systems, namely, seven. Four systems are intimately connected with purposive behavior in individual organisms: (1) consummatory field and riddance patterns, (2) the structures of appetitive and aversive purposive acts, (3) the personal situation, and (4) personality structures.

Pepper acknowledges his profound debt to his long-time associate, E. C. Tolman. Tolman's influence appears clearly in the discussion of purposive behavior. The *consummatory field* includes actions which lead immediately to the quiescence of appetitive drives. *Riddance patterns* have the analogous relation to aversive drives. The selective system defined by these two classes of acts selects affective value, *i.e.* things pleasant or unpleasant. For Pepper, this is the limit of the scope of hedonism in motivation and in value theory. The elements in the structures of purposive acts are *drives* which, initially at least, are cognitively blind, and cognitive "anticipatory sets." Several subordinate acts and related anticipatory sets may be linked in sequence from initiation to quiescence of a drive. In the process of learning many such acts are tried. Some are rejected and others are retained. This selection is in terms of conative-achievement value, otherwise called success or failure. The *personal situation* corresponds very nearly to Lewin's "life-space." It includes various environmental constraints on action. Actions are selected in terms of "prudential" value. Positive and negative values of this type are called prudent or imprudent. As a selective system, *personality structure* faces both toward the environment and inward. The environment-facing function complements the personal situation by determining "what at any given time interests a man, and so forms the pattern of his drives in live-space." Facing inward, personality structure "operates not on acts but on dispositions to act—that is, on drives and their interrelationships" (p. 456). "As a general descriptive definition, we may regard a *personality as a system of dispositions for purposive behavior having its seat in a biological organism*" (p.463, italics in original). Character is the generic term for values thus selected, and its positive and negative aspects are, respectively, integrated and unintegrated.

---

[1] Editor's Note: Professor Diggory recently noted that this book was overlooked at the time of its first appearance, and suggested that it should now be called to the attention of psychologists. His case for the importance of the book was so convincing that we invited him to prepare this delayed notice.

I cannot in such little space do justice to the breadth and facility with which Pepper handles the complex problems of individual purposive acts, but two details must be mentioned. First, he avoids the infinite regress, perennial pitfall of value theorists, by arguing that anything has value only during the process of valuation, a process which occurs only during a current purposive act. Each arousal of an appetitive drive is a separate demand for its satisfaction, and the attainment of its quiescence pattern is separately and *terminally* valued. We need not assume that every goal of every purposive act is necessarily a means to some ulterior end, for the structure of a purpose often shows which acts are terminal and which are subordinate and instrumental. Special mention should also be made of the idea of "value mutation," shifts in the drive-basis for the acceptance or rejection of an act. In his discussion of the "independence mutation" Pepper offers a solution for the so-called functional autonomy problem which is at least as good as any offered previously, and better than most.

When at least two individuals interact to solve a problem a fifth selective system, the *social situation,* comes into play. Selection in a social situation is made according to three criteria: the personal impulses of the participants for purposive satisfaction, demands for conformity, and requirements for the survival of the group. Social values are congenial or uncongenial. Demands for conformity and requirements for survival of the group actually derive from the sixth selective system, *cultural pattern,* as it "legislates over" (determines the criteria of selection in) the social system. Among cultural objects Pepper discusses mainly "functional institutions" such as tribes or nations which define roles for individuals by distributing social functions to them. Cultural values are either conforming or not conforming. They are "legislated over" by the norms of the seventh selective system, *natural selection,* whose values are adaptiveness or unadaptiveness, generically: survival value. Cultural evolution is important for human survival because man can determine the degree of probability of his own survival by the cultural arrangements he invents.

Pepper suggests that his comprehensive hypothesis is open to test in every detail.

In this reviewer's opinion, he is right. I believe he has offered not a finished system but a way of thinking about value and valuing which will be useful for a long time. The kinds of selection he identifies do seem to occur. We may use other terms than he has suggested and we may come better to understand the mechanisms of selection, but we can hardly escape the fact of selection, nor the fact that Pepper's localizations of the various kinds of selection are very nearly correct. People familiar with the literature on value theory will recognize that Pepper's method is not new. In modern times its roots go back at least to Christian von Ehrenfels. More recently, John Dewey and Ralph Barton Perry have continued and improved the tradition, and Pepper acknowledges his debt to both of them. Though Pepper has probably not said the last word about value, he has certainly made the best and most complete statement so far.

University of Pennsylvania JAMES C. DIGGORY

Science in Betting: The Players and the Horses. By E. R. DASILVA and R. M. DORCUS. New York, Harper and Brothers, 1960. Pp. 145. $3.95.

A book written for popular consumption cannot be expected to adhere to the standards of one intended for a professional audience, but *Science in Betting* fails to meet the most modest requirements through an incredible misuse of statistics. Its sales appeal is based on a "scientific approach" to horse-race betting, but it consistently deludes a statistically naïve public with statements which have virtually no statistical reliability and in several instances contradicts factual findings reported in more rigorous treatments.

The authors repeatedly claim that "the public almost always overplays the favorite" (p. 7). Now this comes as quite a surprise, since Griffith and McGlothlin both found that the public underplays the favorite.[1] The latter cited the results of over 9000 races to show the expectation for a $1.00 bet at odds of less than 2.95 to 1 (the usual odds-range of the favorite) to be —$0.07 with a standard deviation of $0.02.[2] Since the track deduction was 13%, an expectation of —$0.13 for a $1.00 bet would indicate neither over- nor under-betting a particular odds-group. Hence, these data show the public *underplays* the favorite and the results lie three standard deviations above the neutral expectation (—$0.13).

In support of their argument that the public overbets the favorite, the authors cite two samples of races in which the odds on the favorite established by the public were 1.82 and 1.99 to 1. The corresponding average odds estimated by the professional handicappers prior to the race were 3.58 and 2.78 to 1. The authors conclude that this demonstrates that the public overplays the favorite. If, however, we use the author's statement that about 30% of all favorites actually win (p. 7), we see that the average expectations for one-dollar bets on the favorites at the handicapper odds will result in the handsome positive values of $0.37 and $0.13 respectively, even after the track deduction. The expectations of one-dollar bets on the favorites by the uninformed "overbetting" public are respectively —$0.15 and —$0.10, and thus rather accurately express the probability of the favorite's winning when the track deduction is considered.

---

[1] R. M. Griffith, Odds adjustments by American horse-race bettors, this JOURNAL, 62, 1949, 290-294; W. H. McGlothlin, Stability of choices among uncertain alternatives, this JOURNAL, 69, 1956, 604-615.

[2] The expectation of a wager is equal to the product of the payoff (odds) and the probability of winning, minus the product of the probability of losing and the size of the bet. The probability of winning for a particular odds-group in horse-racing may be empirically determined by obtaining the proportion of winners in the group for a very large sample of races.

At the conclusion of this extremely naïve argument advanced in support of their erroneous observations that the public overbets the favorite, the authors resume the role of psychologists and offer a "deeper" interpretation for this phenomenon. In their opinion, it brings a "feeling of security" to bet on the favorite (p. 10.). They point out that this security "is not based on fact but on fancy."

At the end of each chapter the authors offer a list of conclusions based on their statistical findings, but with no indication of their reliability. These conclusions are based on grossly inadequate samples throughout the book. For example they state that bets on horses with odds of around 12 to 1 are preferable to those of 6 to 1, "since the percentage of winners is almost the same throughout this range" (p. 26). The very opposite result was found by McGlothlin, using over 9000 races compared to DaSilva and Dorcus' 880 races. The former found the expectation for a $1.00 bet at odds of 6.00-7.95 to 1 was —$0.15, while that for the 8.00-10.95 group was —$0.23. The standard error of the difference between these groups is $0.045, which would indicate that the expectation of the 8.00-10.95 to odds-group is *less* than that of the 6.00-7.95 to 1 group at approximately the 8% level of significance. It is clear that with a standard error of the difference of approximately $0.135 which would result from a sample of 880 races, no conclusions could be drawn regarding which odds-group evidences the higher expectations.

It is perhaps too much to ask DaSilva and Dorcus to provide their readers with an appreciation of statistical significance; however, if they wished to present an accurate picture of their conclusions they might provide the following warning: Should you bet on a number of races equal to the sample we have cited, the chances are about one in three that the results will be opposite to the ones found here.

In conclusion, I should like to raise the question of the effect of such a book on the public image of psychology. It is written by academic psychologists and masquerades under the aura of focusing the powerful tools of science on the field of horse-race betting. For those who critically evaluate the book, it will certainly not enhance the stature of psychology as a science. For that portion of the readership which tends to accept scientifically labeled statements at face value, the authors have failed in their responsibility to provide reasonably accurate information.

WILLIAM H. MCGLOTHLIN

The Rand Corporation

The Motivation of Behavior. By JUDSON S. BROWN. New York, McGraw-Hill, 1961. Pp. xi, 404. $7.50.

Brown has written a good book. It is a detailed critical analysis of the drive concept as Hull originally and Brown more recently have developed it. The concept under scrutiny is the habit multiplier, the energizer, the non-directive, non-associative construct. Dr. Brown has been one of the most constant and ardent proponents of the construct; he has been true to it when many others in the Hullian tradition have gone off to flirt with other, more glamorous constructs. He has lived with the drive concept for a long time, and has come to know both the charms of its promised explanatory power and the harsh realities of its empirical and conceptual application. Both sides of the story are told here with great honesty, objectivity, and with admirable clarity.

The reader is offered two theoretical alternatives: (1) Invoke the drive construct to explain all instances of generalized energization of behavior and use stimuli to explain why one response rather than another becomes energized. (2) Use stimuli to carry the whole explanatory burden. The theorist's choice, then, is whether to distinguish between the activation and the directing of behavior, or whether to explain behavior associatively without invoking any motivational constructs. The second alternative is admitted to have an advantage of parsimony, but, it is argued, it falters empirically on the studies in which hours deprivation or shock intensity is suddenly shifted. There are no widely proclaimed associationistic principles which would lead one to predict that performance should follow appropriately. On the other hand, 'drive' (*D*) handles such cases easily. (Brown's honesty in these matters has led him to formulate an ad hoc associationistic explanation in terms of competing responses that does the job quite nicely. There is some irony in the fact that he has presented a better case for the associationists than any of them has yet done.)

Brown does not distinguish between primary drives (like hunger and shock) and secondary drives (like incentive and fear); he uses one term, drive, to designate the energizing component of both, and regards the old distinction as pertaining to the *source* of drive. Thus, anxiety, frustration, incentive motivation, in short, all of the varieties of "secondary motivation" that have come in recent years to overshadow the primary drives in theoretical importance, are considered to be just learned associative determinants of drive. Incentive motivation, for example, is not incorporated into habit strength, as Hull once had it, or parallel with *D*, as Spence has it, but considered as an S-R determinant of *D*. Brown's *D* thus gains considerable explanatory power over Hull's *D* or Spence's *D*—but it also encounters new difficulties. Indeed, by being made nearly synonymous with motivation in general, *D* acquires nearly all the problems of all motivational concepts. The great promise of the older, narrower *D* was the possibility of anchoring it on the antecedent side; but Brown's *D* is now relatively unanchored antecedently, and so is likely to share with r_G the indignity of being invoked ad hoc.

The treatment of the secondary sources of drive is far less complete, particularly in view of the great complexity of the literature, than the treatment of the primary sources of drive. The secondary sources of drive are acknowledged to follow associative principles and to be elicited by stimuli. So we get the feeling that the drive theorist has far less of an advantage over the association here. The score is getting evened up. A lot of theorists have become reconciled to this; they don't care too much, for example, if r_G and fear are drive-like or stimulus-like as long as they can be tied systematically to experimental variables. But Brown, by sharpening, and insisting upon, the distinction between the energizing and the directing of behavior, is forcing a showdown. We begin to see that the fundamental question for motivation theory is: How can we tell when a stimulus motivates behavior and when it merely elicits behavior?

A new and promising research area is suggested by Brown's treatment of perception. Perceptions are regarded as logical constructs which have some of the same properties as responses, but which have the special property that they make possible delayed response to a stimulus after the stimulus has been removed. Ways are discussed of testing the effect of drive on these intervening, response-like perceptions, but, so far, disappointingly little research has been developed with these new conceptual tools. After considering much of the literature on perception as a function of motivation (the *New Look*), Brown concludes that on this question the drive theorist has little or no advantage over the associationist.

The book is remarkably free from errors. It is certainly suitable as a text in a motivation course with a strong orientation toward behavioral theory, but its silence on matters Freudian and Lewinian will probably keep it from being too widely used.

Hollins College

ROBERT C. BOLLES

Psychology of Personality. HUBERT BONNER. New York, The Ronald Press Company, 1961. Pp. vi, 543. $7.00.

Bonner's new text is an attempt at a comprehensive coverage of the personality field from the *holistic* point of view. He makes it clear at the outset that "while not rejecting the mechanistic approach entirely, [he believes] . . . that it has been largely outgrown in the physical sciences, and . . . there is no convincing reason why it will not in brief time be outgrown in psychology itself" (p. 4).

The book is a large one, systematically organized (though with considerable redundancy), and holds true to its bias from preface to conclusion. Supporters of the views of Gordon Allport, Kurt Goldstein, Kurt Lewin, A. H. Maslow, Gardner Murphy, Carl Rogers, and other theorists of self-actualizing, emergent human personology will find themselves at home here, and some may feel that Bonner's attempt to integrate these theories has more to offer as a text than the side-by-side treatment of Hall and Lindzey (1957). On the other hand, hard-core behaviorists (and even some soft-core behaviorists) will undoubtedly react negatively both to the admitted one-sidedness of the author's treatment and to his selection of materials to support his views.

The book is divided into four parts. Part I, an overview, contains a chapter on "What Is Personality?", two chapters on theories—"Points of View" and "Typologies"—, and one on "How Personality is Studied." Part II deals with determinatns of the formation of personality, and is divided into chapters on "Organic Foundations," "The Human Group," and "Enculturation." Part III (which includes over half of the book) is a pot pourri on organization and dynamics: "Psychological Analysis of Personality." It has nine chapters on topics of learned behavior, motivation, culture and purposive behavior, language, expressive behavior, attitudes and values, character and moral behavior, and "The Total Personality" (treated in separate chapters on its structure and the self). Part IV is a single chapter containing a summary, interpretation, conclusions, and an approach to integration of the (holistic) points of view.

Footnoted references are extensive, but are heavily weighted with social psychological sources—not unexpected since Bonner has recently authored two books in that area. A number of references are of recent date, but in several instances recent material from sources cited is overlooked (*e.g.* Bonner disagrees with a number of points raised by Stagner in the 1948 edition of his book on personality. Stagner has revised and extended some of these views since, and now has a new edition of this book out.) This will limit the usefulness of his integration in so rapidly changing a field.

Likewise, one seeks in vain references to such important work as that of Asch in the discussion of conformity, that of Ashby, Wiener, Stagner, Flügel, or Helson in the discussions of homeostasis (which incidentally is treated several places in the book but fails to make the index), that of the ethologists in the discussions of instinct, that of Lindsley or Magoun or Selye in the discussions of the nervous or endocrine systems (the latest reference cited is 1951), or to Mowrer's provocative ideas on guilt in the discussions of moral behavior, or to Rappaport, Colby, Hartmann, Kris or Erikson in discussing new psychoanalytic thinking about the ego. Admittedly, the print explosion in psychology puts an author of a textbook at a disadvantage, but it is our opinion that too many significant trends (and the above are but some examples) have been neglected to warrant nomination of this particular work as *the* definitive integration.

A further criticism, from this reviewer's point of view, is the heavy redundancy in the book. There is no question that Professor Bonner is on one side of a fence—he clearly informs us of this at the outset—but he returns to his argument again and again and again. Surely the selection of this book as a course text would presume a sympathetic instructor.

There is also considerable repetition of specific arguments, such as the one that the various theories of self have much in common. This redundancy could be omitted and the space saved used either to broaden the scope of coverage (*e.g.* Bonner expresses the regret in a footnote that he does not more extensively treat Roger's self theory), or to support some of the positions described by presenting more experimental data (*e.g.* Child's research on somatotypes is not mentioned although more than three pages are spent evaluating Sheldon's work).

Finally, to put the above in perspective, let us say that this is a valuable attempt at synthesis of parallel points of view, and should thus provide a useful source book on self-theories of personality. It is to be regretted only that the author did not choose the more limited title "*Holistic approach to the psychology of personality*" which would more accurately reflect the book's content. If this were its title, this reviewer for one would have taken a much more positive view of what is here offered.

Southern Illinois University MORTIMER H. APPLEZWEIG

Teaching and Learning in Medical School. Edited by G. E. MILLER, with STEPHEN ABRAMSON, I. S. COHEN, H. P. GRASER, R. S. HARNACK and ADELLE LAND. Cambridge, Mass., Harvard University Press, published for The Commonwealth Fund, 1961. Pp. xiii, 304. $5.50.

It is not often that the professional school becomes the focal interest of the professional educator. Rather, educators have been largely concerned with the primary, secondary and collegiate educational process. Recently, self-conscious efforts have been made to elucidate the manner in which young college graduates are transformed into professional practitioners. This book is the outgrowth of one such effort.

Medical schools in the United States are beginning to ask themselves fundamental questions about the professional training of the physician. Who applies to medical school? How is admission achieved? What are the principles of curriculum formation? What criteria for curriculum change are employed? Are medical educational goals being achieved or frustrated? and are such goals adequate and defensible?

The project in Medical Education was a five year program at the University of Buffalo School of Medicine which emerged from the concern of medical educators to seek ways of understanding the educational process—teaching and learning—in medical school. The project eventually combined the interests of the medical school and the school of education—medical educators and professional educators.

The book itself is intended as a source book and reference work for medical educators. It seeks to shed light on the major variables affecting educational success in medical school: the selection of students, the initial adjustment to medical education, the phenomena of learning, principles and techniques of teaching, instructional aides and the evaluation of the instructional effort. What emerges is a sensible, eclectic, practical and extremely readable guide to those who would teach medical students. It shows none of the doctrinaire approaches to either medicine or education. It holds no special brief for any one 'system' of either learning or teaching. The authors avow debts to the theoretical thinking of both the associationist and Gestalt viewpoints. They avoid espousal of any theoretical bent since they believe that "most learning psychologists are interested in advancing psychology as a science, not in making and direct contribution to the improvement of education. The extrapolation of their studies to the classroom may really be unjustified or premature" (p. 48). Nevertheless, faced with the necessity of founding teaching and learning upon a theoretical base, both Thorndike and Lewin are invoked. Such constructs as the meaningful context, motivation, overlearning, transfer, and feedback are alluded to. In the reviewer's opinion, systematic associationism comes off second best as the foundation upon which 'principles of learning' are offered.

A central objective of this book is to point out a variety of instructional techniques appropriate for medical education. Readers may be offended by some of the simplicistic advice given (*i.e.* "Cultivate a good time-sense or use a timepiece"; "do not monopolize . . . discussion"; . . . "let the audience know the objectives of the lecture").

But there is much of value to be found by the discerning medical educator. The new instructor will find here a short course in basic instructional tools. Others will be interested in comparing the objectives of education reviewed with their own principles and values. On one level the book is a thoughtful provocateur; on another it is a primer in method. The latter even includes a discussion of techniques for evaluating knowledge, attitudes, and performance. At the very least the effect will augment any latent beliefs that such phenomena are measurable, let alone worth measuring.

The collaborative efforts of behavioral scientists, educators, and medical educators and practitioners are only in their infancy. This book is a worthy effort and should inspire further examination of the many potential areas of mutual concern.

BERNARD KUTNER

Albert Einstein College of Medicine

The Birth of the Gods: The Origin of Primitive Beliefs. By GUY E. SWANSON. Ann Arbor, the University of Michigan Press, 1960. 260 pp. $4.95.

Guy Swanson has put on his sociologist's hat, picked up a well-thumbed copy of Durkheim's thoughts on the relation of the supernatural beliefs of a people and their soical organization, borrowed the 'cross-cultural method' from the anthropologist Murdock, and, after a good deal of scholarly cogitation, written and tested some intriguing hypotheses about the origins of religious beliefs among simple peoples. This is no small order, and it is pleasant to welcome a first-rate sociological mind to an endeavour previously largely restricted to a few anthropologists and psychologists. The research art has been advanced, science is served, and some mystery has been removed from the possible origins of seven ideas: the conceptions of a monotheistic deity, of polytheistic gods, of ancestral spirits, reincarnation, the immanence of the soul, the prevalence of witchcraft, and the notion of gods who concern themselves with human moral problems.

The core idea is Durkheim's, but it is elaborated and rendered testable by Swanson: these supernatural ideas reflect certain particular naturalistic conditions of social organization in societies as these conditions are interpreted by human experience. The origins are of the supernatural ideas—social origins, in this view. Consider monotheism. This should arise where man has had experience with the mysterious workings of a type which is superordinate over other types to which the first brings order. This higher group needs to have "at least two such subordinates at hand in order to create a relationship between them." Therefore Swanson tests the hypothesis that a conception of a high god will require the presence of at least three types of groups arranged in a hierarchy. His data are ratings of the contents of 50 extant ethnographies, one from each of Murdock's world culture areas (more or less) with proven independent rater reliability. His findings on monotheism are impressive: of the 199 societies having fewer than three sovereign groups, 17 lack monotheism. Of the 20 societies having three or more such groups, 17 possess a high god. The probability of so great a difference occurring by chance is less than 0.0005, by a one-tailed chi-squared test. Alternative interpretations are considered, and a number of important and strongly held notions fall by the wayside against these empirical criteria.

Swanson has good luck with most of his hypotheses, but writes about them in a reasonable, restrained manner which is in contrast to the boldness of the book's title. Swanson stands firm on the idea that in playing off his notions about 'origins' of supernatural beliefs against most of the reasonable or important alternatives, he is finding out about origins as much as anyone can do so. He is boldly and refreshingly causal in his mode of thinking and is aware of the need for data on culture change to provide additional tests of his hypotheses.

This book is pleasant to read if one does not carry too much of a chip on one's shoulder against Durkheim. It is a chastening experience for one who has a bias against the assumption of a 'group mind'; or groups as 'embodying purposes.' Swanson shows that this kind of thinking can order data, and that it can suggest new orderings that psychological thinking has not as yet considered.

It is a shame that Swanson does not worry more about geographical distributions in his findings that might suggest diffusion instead of his 'independent emergence' kind of explanation, but the raw data are included for doubters of this kind. Swanson was also not aware of other data of a similar kind that has been developed, but this should not keep us from crediting him with a tentative success in the search for the beginnings of a naturalistic explanation of supernatural notions.

Cornell University
W. W. Lambert

Principles of Comparative Psychology. Edited by Rolland H. Waters, D. A. Rethlingshafer, and Willard E. Caldwell. New York, McGraw-Hill Book Company, Inc. 1960. pp. ix-453. $7.95.

This volume comes in the traditional format set by Moss in 1934 and continued through the last edition of Stone's *Comparative Psychology* in 1951. It is essentially a collection of chapters written by specialists in the field of animal behavior. Although most of the contributors attempt to approach their subject matter from a broad comparative view, the treatment of information outside of the traditional laboratory animals (a limited number of mammals), or outside the author's speciality, is either limited or shallow. In this respect, the editors have not departed from tradition.

One of the significant deviations from the older books is the reduction of emphasis on learning and learning theory. While Moss (1942) devoted 43% of the total volume to these topics and Stone (1951) 33%, this text devotes only 16%. This change probably reflects a continuous change of interest on the part of American compartive psychologists over the years. Innovations of interest are a comprehensive chapter dealing exclusively with acquired drives and curiosity motives and individual chapters on the problem of classification, the theoretical foundations of comparative psychology and research trends in comparataive psychology.

The latter three chapters provide a degree of unity and summarization that is usually not found in works of this type. They also produce, however, one of the major shortcomings of the book, redundancy and needless repetition of subject matter. Many of the individual chapters discuss one or all three of these topics to a considerable degree. This weakness is not the fault of the individual contributors, but of the editors.

Omission and selection of material are unavoidable in any survey text, but on some points this book is excessive. For example, little or no mention is made of the use of operant techniques in the study of sensory processes, secondary reinforcement, learning and complex processes. The chapter on learning devotes four paragraphs (less than two pages) to instrumental learning and thirteen pages to *selected* experiments in support of an 'integration' theory of learning. The chapter on sensory processes omits the elegant experiments by Guttman and Blough. More than one-fourth of the chapter on social behavior is concerned with a detailed analysis of the socialization of sheep and the application of the behavioral categories developed with this animal to comparative social behavior.

Although the book has its weaknesses, it should be welcomed by the teacher and student of comparative psychology. It should serve as an excellent start in the field for the advanced undergraduate and the beginning graduate student, but if used in isolation might tend to give the student a narrow, if not biased, point of view in many areas.

Robert R. Zimmermann

Cornell University

L'Automation; Aspects Psychologiques et Sociaux. By ALPHONSE CHAPANIS, A. LUCAS, EUGENE JACOBSON, NORMAN MACKWORTH, LEONARDO ANCONA and GUSTAVO IACONO. Louvain, Publications Universitaires de Louvain, 1960. Pp. 119. 11NF.

The papers in this monograph are modified from a series originally presented in a symposium organized by Mackworth for the Fifteenth International Congress of Psychologists in Brussels, August, 1957. Much of the material is available elsewhere, particularly in engineering and military publications, but this symposium presents a summary and integration of previously reported work. Some new content is included in most papers. Several topics usually considered are scarcely mentioned (*e.g.* the problems involved in vigilance, or individual variables related to performance of automated tasks). Nevertheless, the summary will be convenient for industrial and social psychologists with reasonable facility in reading French.

Chapanis ("Le Facteur Humain dans la Construction des Systèmes") presents a survey of the literature and some clear recommendations for design, together with examples. Lucas, of Régie Renault ("L'Automatisme dans l'Industrie Automobile") describes in detail numerous automated operations, discussing the actual changes occurring with automation. Jacobson ("Quelques Aspects Socio-Psychologiques des Reactions de l'Employé devant l'Évolution Technique") outlines a program for research on adaptation to change, surveying a number of studies including some newly completed or then in progress. Mackworth ("Automation et Méthodes de Recherche") sketches the characteristics of automated work. Ancona and Iacono ("Workers' Attitudes to Technological Change: Some Observations") summarize (in English) a study of the impact of technological change upon a group of workers in a steel mill.

Symposia tend to be both redundant and somewhat disjointed; this is no exception. Several themes recur, however, beyond the obvious overlapping of bibliographies.

(1) Technological change is usually a matter of evolution rather than of revolution.

(2) The effects of automation on tasks and situations vary greatly, although working conditions are likely to be improved.

(3) The human operator continues to be important even in so-called "automated" tasks.

(4) Workers have accepted automation when intelligently introduced.

(5) Research is needed concerning design of tasks, optimal use of men and machines in systems, and the factors influencing adaptation to change.

The problems discussed will suggest numerous challenging topics for systematic research.

Cornell University

PATRICIA C. SMITH

Markov Learning Models for Multiperson Interactions. By PATRICK SUPPES and RICHARD C. ATKINSON. Stanford, Stanford University Press, 1960. Pp. xii, 296. $8.25.

The authors of this book describe a series of experiments in which the principal interest is in the degree of which stimulus-sampling theory and game theory predict the behavior of subjects "interacting" in a competitive game-like situation. The statistical learning theory does the better job, but this is not too surprising, since game theory is an appropriate model for a single trial, whereas learning theory is a more appropriate model for a series of trials.

Suppes and Atkinson apply an all-or-nothing version of stimulus sampling theory to a basic game in which each player has a choice of two responses on each of a number of trials, and on each trial receives or does not receive approval. The conditions that are varied over the experiments include the subjects' knowledge of, and the facts of, the relation between his responses, an opponent's responses, and his rewards. Although the responses obtained in the experiments do not represent independent sequences, they may in some instances be described adequately by first-order Markov chains. The one-element model predicts over-all levels of performance in a wide variety of experiments best, but does not do an adequate job for the detailed sequential responses. A two-element model does better for predicting the sequences but not so well for mean response-level. Even though the analyses are made with the aid of a computer, they rapidly become complex and unwieldy for groups larger than two. A "generalized conditioning model" is introduced in the book and is considered to be "promising," but it is so complex that fitting it to the data of some of the experiments is not attempted.

The book may be considered a progress report of a large-scale research project. More than a 1000 subjects were used in the experiments. In addition to the reports of experiments, the authors provide background material and discussion of statistical learning models and statistical methods used in estimating parameters and testing goodness to fit, including a valuable presentation of Markov (especially one-stimulus element) models. There is also some discussion of how stimulus-sampling theory might be applied to the study of social conformity, economic oligopoly, and a section on continous games. We may look forward to further reports of experiments in these and related areas.

The orientation of the authors is mathematical, which will make the book difficult for some readers interested in the subject matter. Those who have the necessary background, or who are willing to work through it, will find portions of this original contribution rewarding.

The University of Texas
BENJAMIN FRUCHTER

Electrical Stimulation of the Brain: An Interdisciplinary Survey of Neurobehavioral Integrative Systems. Edited by DANIEL E. SHEER. The University of Texas Press, 1961. Pp. xvi, 641. $17.50.

This book is important; it is also large. Important books are not rare in the broad field of physiological psychology but generally their importance is limited by their scope. Most are written by two or three authors who tend, quite rightly, to emphasize their own work or to write books of the "introductory type." In either case something gets slighted. This book, to the contrary, covers the entire field with great fullness. It was authored by many (55!) of the most important men in its area. Size and scope are the assets and liabilities of this book, but the assets are many, the liabilities few.

Not only are history, experimental results, and interpretations reported here, but 13 of the chapters give detailed expositions of the techniques employed which are well illustrated with drawings and photographs. The title might well be "How To Do Your Own Electrical Stimulation of the Brain." The book contains 40 chapters divided into 7 sections, and one chapter of stereotaxic atlases of the rat, cat, and monkey. An indication of the book's completeness is in the titles of the sections: Introduction, Techniques, Neuroanatomical Correlates, Neurophysiological Correlates, Behavioral Correlates, Experiential Correlates, and Summary. The book closes with a Bibliography, a Name- and a Subject-matter Index.

This book might be expected to suffer from so many different contributors having differing styles and aptitudes at writing but it does not. Dr. Sheer, its editor, has given us a multi-authored book that is remarkably uniform in its writing.

The liabilities, as has been said, are few, and these are not serious. The book should be very useful, particularly in the laboratory though its sheer bulk might make this difficult. One wishes, for instance, that the stereotaxic atlases could be printed separately.

The Hogg Foundation for Mental Health deserves commendation for supporting the publication of this book. *Electrical Stimulation of the Brain* will be relied on as authoritative for many years to come.

University of Texas BRUCE H. DEATHERAGE

Personality Dynamics and Effective Behavior. By JAMES C. COLEMAN with selected readings prepared by ALVIN MARKS. Chicago, Scott, Foresman, 1960. Pp. 566. $5.50.

This is a textbook intended for college classes, concerned with the psychology of adjustment. Author Coleman takes as his theme our world of technology and its ensuing problems of human adjustment in times of global upheaval, social unrest, and near-realized hopes for a golden age. His emphasis is on the adjustment of the normal person with only a short detour into the realm of psychotic states. This last is a refreshing contrast to the space allocations of most texts which purport to deal with the normal personality but which actually devote half or more of their available pages to the severe behavior disorders. Coleman's book sticks closely to the subject matter appropriate to the title, and the material is presented in one of the clearest expository styles this side of William James. The organization of the volume is that of the traditional reference frame of psychology—learning, emotion, perception, etc.—as applied to the adjustment process. There is one brief excursion from research data, a leap into a murky cavern of what Coleman calls "assumptions concerning value." But if the usual landmarks of hard evidence are missing in this portion, the excursion he makes *is* brief and, in a way, serves to point up Coleman's more usual hard-shell approach.

An unusual feature of the book is a closing section on "Explorations on Personality and Behavior Theory" prepared by Alvin Marks. This concluding section is addressed to the more serious undergraduate student of psychology and consists of selected readings from the writings of Gordon Allport, J. P. Guilford, Nicholas Hobbs, Melvin Marks, James G. Miller, James Olds, Carl Rogers, and others. The readings represent a sampler of what is going on in such topical areas as psychological theory construction, cybernetics, and ethical behavior. The selections are well chosen and should prove challenging to the college student.

Louisiana State University IRWIN A. BERG

Challenge of Psychical Research: A Primer of Parapsychology. By GARDNER MURPHY. New York, Harper and Brothers, 1961. Pp. xviii, 297. $6.00.

Dr. Murphy, Director of Research at the Menninger Foundation, writes about his life-long interest in psychical research (telepathy, clairvoyance, precognition, psychokinesis, and survival after death) for a series entitled "World Perspectives," the purpose of which is "to reveal basic new trends in modern civilization, to interpret the creative forces at work today. . . ." The book "is aimed at two kinds of people: the professional scientist and the thoughtful lay reader."

In the opinion of this reviewer, the book does not hit either target. For the professional scientist, there is inadequate consideration of the criticism of scientific methodology, particularly the role of the experimenter in recording and reporting objective events, that have been offered during the past eighty years. For the thoughtful lay reader, there is too much "scientificating" with tables of data, probability statistics and description of minutia.

Murphy puts the problem for the professional scientist in the proverbial nutshell on p. 100: "There is no place at Duke University, or City College, or Harvard, or the Society for Psychical Research in London, where time and place can be set for a demonstration in which scientists can find all their difficulties openly, simply and fully met. Hand in hand with this is the fact that the investigator cannot describe in full the conditions under which he would certainly obtain such results in his own laboratory. The inquiring scientist is told that something wonderful has happened and been observed by people whom he does not know. He is not in the habit of believing on this basis, especially when the results run counter to his major assumptions about reality."

Princeton University JOHN L. KENNEDY

The Philosophical Roots of Scientific Psychology. By HENRYK MISIAK. New York, Fordham University Press, 1961. Pp. 142. $3.00.

This is an excellent little book, which students of the history of psychology will find useful. It is obviously not a substitute for Boring, and certainly not for Brett; but with amazing skill it packs into very small compass the essentials of psychology's philosophical heritage. It reminds one of the *repetitur* of the European university of an earlier generation, a clear and compact statement of the minimum that the candidate will be expected to remember. This is not intended as a disparagement; a *repetitur* can be an important adjunct to education, and this is a good one.

Father Misiak came to us from Poland, where he presumably received his early training in Scholastic psychology. He paused for a time in Scotland, where he must have polished his English—for his English is now impeccable. While in Scotland he must have learned something about parsimony, for he knows how to boil down important thoughts into a very few words. He is now at Fordham, and he has already given us, with the collaboration of Virginia M. Staudt, an important volume on *Catholics in Psychology: A Historical Survey,* 1954.

This reviewer, not a Roman Catholic, confesses that he expected a Roman bias. He is delighted to report that he found nothing of the sort. Father Misiak is completely open in his approach to history and fully knowledgeable in his approach to psychology. The book is terribly compressed, of course, but the clarity and balance deserve the highest commendation. The reader, and this means the teacher as well as the student, will profit from the reminder that the problems of modern psychology gain in intelligibility when they are viewed against the background of their history.

Cornell University ROBERT B. MACLEOD

Elements of Mathematical Statistics. By HOWARD ALEXANDER. New York, John Wiley & Sons, 1961, Pp. xi, 367, $7.95.

This is an elementary book of mathematical statistics which treats traditional topics in a classical manner. It heavily emphasizes inferential statistics, especially sampling distributions, in contrast to descriptive statististics. An introduction to the theory of sets and matrices is included. Since the book requires a two semester course in calculus, it would not be appropriate for use as a textbook in most psychological statistics courses. For the person who has had calculus, the *Elements of Mathematical Statistics* may serve its most useful function in providing a clear introduction to the terms and concepts necessary for comprehension of a more sophisticated statistical literature.

Cornell University JASON MILLMAN

Psychotherapy with Schizophrenics: A Reappraisal. Edited by J. G. DAWSON, H. K. STONE, and N. P. DELLIS, introduction by JOHN N. ROSEN. Baton Rouge, Louisiana State University Press, 1961. Pp. xii, 156. $5.00.

The purpose of this book, a report of a symposium held at Southeast Louisiana Hospital, was to consider the methods and theory in the treatment of schizophrenics developed since the publication of Brody and Redlich's book in 1952. The major contributors to the current volume are Carl Rogers, Russell Monroe, Eugene Brody, Albert Scheflen, Robert Roessler, and Thomas Malone. A formal discussion follows each paper.

If the purpose of the book is to examine the methods and theory of treatment, the book fails to do so adequately. Each contributor is narrow in his presentation and fails to enunciate a clear plan of procedure. If the purpose of this volume is to "be regarded as a sampling of thought and attitude" about schizophrenia by therapists then an evaluation becomes impossible.

After reading this book one is left with the following impressions: (1) not much has happened since the Yale symposium some 10 years earlier; and (2) not all symposia should be published.

University of Minnesota Medical School REUBEN J. SILVER

BOOKS RECEIVED

A continuation of the books listed in the September number (Vol. 75, 1962, 525 ff.) Listing here does not preclude their later review.

MICHAL-SMITH, HAROLD, and KASTEIN, SHULAMITH, *The Special Child,* Seattle, Wash., Bureau of Publications, 1962, pp. 350, $5.50.

MOK, PAUL P., *A View From Within,* New York, Carlton Press, Inc., 1962, pp. 182, $3.50.

MORGENSTERN, IRVIN, *The Dimensional Structure of Time,* New York, Philosophical Library, Inc., 1960, pp. 174, $3.75.

MÜLLER, W. H., and ENSKAT, A., *Graphologische Diagnostik,* Bern und Stuttgart, Verlag Hans Huber, 1961, pp. 304, Fr/DM 44.

MYERS, F. W. H. (ed. by Susy Smith), *Human Personality and its Survival of Bodily Death,* New Hyde Park, N.Y., University Books, Inc., 1961, pp. 416, $10.00.

NASH, HARVEY, *Alcohol and Caffeine,* Springfield, Ill., Charles C Thomas Co., 1962, pp. xi, 169, $7.25.

NOSOW, SIGMUND and FORM, WILLIAM (eds.), *Man, Work and Society,* New York, Basic Books, Inc., 1962, pp. viii, 612, $8.50.

OPPENHEIM, FELIX, *Dimensions of Freedom,* New York, St., Martin's Press, 1961, pp. 242, $6.50.

OSBORN, ARTHUR, *The Future is Now,* New Hyde Park, New York, University Books, Inc., 1962, pp. 254, $6.00.

PARRATT, LYMAN, *Probability and Experimental Errors in Science,* New York, John Wiley and Sons, 1961, pp. xv, 255, $7.25.

PATTERSON, C. H., *Counseling and Guidance in Schools,* New York, Harper and Bros., 1962, pp. xiii, 382, $4.75.

PATTON, ROBERT *et al., Current Trends in Psychological Theory,* Pittsburgh, Pa., University of Pittsburgh Press, 1961, pp. ix, 229, $6.00.

PICKARD, P. M., *I Could a Tale Unfold: Violence, Horror and Sensationalism in Stories for Children,* New York, Humanities Press, Inc., 1961, pp. xi, 228, $5.00.

PIRE, G., *Tests Ikonia A; Tests Ikonia B; Manuel d'Instructions, Tests Ikonia A, p. 68; Manuel d'Instructions, Tests Ikonia B,* p. 55, Paris, H. Dessain.

PITTENGER, ROBERT E., HOCKETT, CHARLES F., and DANHEY, JOHN J., *The First Five Minutes. A Sample of Microscopic Interview Analysis,* Ithaca, N.Y., Paul Martineau, 1960, pp. ix, 264, $6.50.

POLLINI, FRANCIS, *Night.* Boston, Mass., Houghton Mifflin Co., 1961, pp. 305, $3.95.

POSTMAN, NEIL, *Television and the Teaching of English,* New York, Appleton-Century-Crofts, Inc., 1961, pp. vii, 138, $1.00.

RADZINOWICZ, LEON, *In Search of Criminology,* Cambridge, Mass., Harvard University Press, 1962, pp. vii, 254, $4.75.

REIK, THEODOR, *The Temptation,* New York, George Braziller, 1961, pp. 256, $4.50.

REISS, ALBERT J., *Occupations and Social Status,* New York, The Free Press, 1962, pp. vii, 305, $5.00.

RIDENOUR, NINA, *Mental Health in the United States,* Cambridge, Mass., Harvard University Press, 1961, pp. 146, $3.50 (A Commonwealth Fund Book).

ROBACK, A. A., *History of Psychology and Psychiatry,* New York, Philosophical Library, Inc., 1961, pp. 422, $7.50 (cloth), $2.25 (paper).

ROSENBLATT, FRANK, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms,* Washington, D.C., Spartan Books, 1962, pp. xvi, 616, $6.50.

ROYCE, JAMES E., *Man and His Nature,* New York, McGraw-Hill Book Co., Inc., 1961, pp. xii, 398, $5.50.

RUNES, DAGOBERT, *The Art of Thinking,* New York, Philosophical Library, Inc., 1961, pp. 90, $2.75.

SADER, MANFRED, *Möglichkeiten und Grenzen Psychologischer Testverfahren,* Bern und Stuttgart, Verlag Hans Huber, 1961, pp. 107, Fr/DM 9.80.

SALZMAN, LEON, and MASSERMAN, JULES H., *Modern Concepts of Psychoanalysis,* New York, Philosophical Library, Inc., 1962, pp. 210, $4.75.

SARBIN, THEODORE, *Studies in Behavior Pathology,* New York, Holt, Rinehart and Winston, 1961, pp. x, 341, $4.00.

SELLS, S. B., *Essentials of Psychology,* New York, Ronald Press, Inc., 1962, pp. xvi, 513, $6.50.

SHIPLEY, JOSEPH, *The Mentally Disturbed Teacher,* Philadelphia, Pa., Chilton Books, 1961, pp. vii, 404, $5.00.

SODDY, KENNETH, *Clinical Child Psychiatry,* London, Bailliere, Tindall and Cox, 1960, pp. x, 470, $8.50.

SPITZ, RENE A., *Genetic Field Theory of Ego Formation,* New York, International Universities Press, 1960, pp. 123, $3.00.

STENGEL, E., *Classification of Mental Disorders,* Geneva, World Health Organization, 1960, *21,* pp. 601-663, 60¢.

STROTHER, GEORGE (ed.), *Social Science Approaches to Business Behavior,* Homewood, Ill., Richard D. Irwin, Inc., 1962, pp. vii, 183, $5.00.

SUMMERS, MONTAGUE, *The Vampire in Europe,* New Hyde Park, N.Y., University Books, Inc., 1962, pp. xx, 329, $7.50.

SZONDI, L. et al., *Beiträge zur Diagnostik, Prognostik und Therapie des Schicksals,* Bern und Stuttgart, Verlag Hans Huber, 1962, pp. 280.

THOMA, HELMUT, *Anorexia Nervosa,* Bern, Switzerland, Hans Huber Publishers, 1962, pp. 352, Fr/DM 18.50.

THOMPSON, LAURA, *Toward a Science of Mankind,* New York, McGraw-Hill Book Co., 1961, pp. xxvii, 276, $5.95.

THORPE, W. H., and ZANGWILL, O. L., eds., *Current Problems in Animal Behaviour,* Cambridge, Eng., Cambridge University Press, 1961, pp. xiv, 424, $8.50.

THORWALD, JURGEN, *The Dismissal*, New York, Pantheon Books, Inc., 1962, pp. 256, $5.50.

TRAPP, E. PHILIP, and HIMELSTEIN, PHILIP (eds.), *Readings on the Exceptional Child,* New York, Appleton-Century-Crofts, Inc., 1962, pp. xii, 674, $7.00.

TUSSING, LYLE, *Study and Succeed,* New York, John Wiley and Sons, 1962, pp. ix, 157, $2.95.

TYRRELL, G. N. M., *Science and Psychical Phenomena,* and *Apparitions,* New Hyde Park, N.Y., University Books, Inc., 1961, pp. ix, 379; 172, $7.50 (2 books bound together in one volume).

VIAUD, GASTON, *Intelligence: Its Evolution and Forms,* New York, Harper and Bros., 1960, pp. 127, $1.75.

VITELES, MORRIS S., BRAYFIELD, ARTHUR, and TYLER, LEONA E., *Vocational Counseling,* Mineapolis, Minn., University of Minnesota Press, 1962, pp. 70, $1.50.

VON ABELE, RUDOLPH, *The Vigil of Emmeline Gore,* Boston, Mass., Houghton Mifflin Co., 1962, pp. 263, $4.00.

VON BÉKÉSY, GEORG, *Experiments in Hearing,* New York, McGraw-Hill Book Co., 1960, pp. x, 745, $25.00.

WALKER, KENNETH, *The Extra Sensory Mind*, New York, Emerson Books, Inc., 1961, pp. 256, $3.95.

WEBB, WILSE, ed., *The Profession of Psychology,* New York, Holt, Rinehart and Winston, Inc., 1962, pp. viii, 291, $4.50.

WECHSLER, LOUIS, BLUM, MARTIN, and FRIEDMAN, SIDNEY, *College Entrance Examinations,* New York, Barnes and Noble, Inc., 1960, pp. 305, $1.95.

WEDECK, HARRY, *Dictionary of Aphrodisiacs,* New York, Philosophical Library, Inc., 1961, pp. 256, $10.00.

WEINSTEIN, EDWIN A., *Cultural Aspects of Delusion,* New York, The Free Press, 1962, pp. 215, $5.00.

WESTOFF, CHARLES F., POTTER, ROBERT G., JR., SAGI, PHILIP C., and MISHLER, ELLIOT, G., *Family Growth in Metropolitan America,* Princeton, N.J., Princeton University Press, 1961, pp. xxii, 433, $10.00.

WILKS, SAMUEL, *Mathematical Statistics,* New York, John Wiley and Sons, 1962, pp. xvi, 544, $15.00.

WINN, RALPH (trans. and ed.), *Psychotherapy in the Soviet Union,* New York, Philosophical Library, Inc., 1961, pp. 207, $6.00.

WITKIN, H. A., DYK, R. B., PATERSON, H. F., GOODENOUGH, D. R., and KARP, S. A., *Psychological Differentiation,* New York, John Wiley and Sons, 1962, pp. xii, 418, $7.95.

YACORZYNSKI, G. K., ATKINSON, ARTHUR, COHEN, JEROME, and SHUFFLEBARGER, F. G., *Investigation of Carbon Dioxide Therapy,* Springfield, Ill., Charles C Thomas, 1962, pp. viii, 313, $9.75.

REVISED EDITIONS

ANASTASI, ANNE, *Psychological Testing,* 2nd ed., New York, Macmillan Co., 1961, pp. xiii, 657, $7.50.

CALHOUN, DON, NAFTALIN, ARTHUR, NELSON, BENJAMIN N., PAPANDREAU, ANDREAS, G., and SIBLEY, MULFORD Q., *An Introduction to Social Science,* 3rd ed., Philadelphia, Pa., J. B. Lippincott Co., 1961, pp. xix, 392, 361, 370, xxxi.

FESTINGER, LEON, *A Theory of Cognitive Dissonance* (Reissue), Stanford, Calif., Stanford University Press, 1962, pp. xi, 291, $5.25.

GARRETT, HENRY E. (assisted by Bonner, Hubert), *General Psychology,* 2nd ed., New York, American Book Co., 1961, pp. xii, 649, $6.75.

LAIRD, DONALD A., and LAIRD, ELEANOR, *Practical Business Psychology,* 3rd ed., New York, McGraw-Hill Book Co., 1961, pp. iv, 442.

LINDGREN, HENRY CLAY, *Educational Psychology in the Classroom,* 2nd ed., New York, John Wiley and Sons, 1962, pp. xvii, 574, $6.95.

MADSEN, KRISTEN BENT, *Theories of Motivation,* 2nd ed., Cleveland, Ohio, Howard Allen, Inc., 1962, pp. 356, $7.00.

MATHENEY, RUTH V., and TOPALIS, MARY, *Psychiatric Nursing,* 3rd ed., St. Louis, Mo., C. V. Mosby Co., 1961, pp. 281, $3.75.

MORGAN, CLIFFORD, *Introduction to Psychology,* 2nd ed., New York, McGraw-Hill Book Co., 1961, pp. xvii, 727, $7.50.

ROHRACHER, HUBERT, *Kleine Charakterkunde,* Neunte Auflage, Wien and Innsbruck, Urban and Schwarzenberg, 1961, DM 14.-ö.S. 89.

SCHNEIDERS, ALEXANDER A., *Personality Development and Adjustment in Adolescence* (rev. ed. of *Psychology of Adolescence*), Milwaukee, Wisc., The Bruce Publishing Co., 1960, pp. x, 473, $5.75.

SCHONELL, FRED J., *The Psychology and Teaching of Reading,* 4th ed., New York, Philosophical Library, Inc., 1961, pp. 295, $6.00.

STAGNER, ROSS, *Psychology of Personality,* 3rd ed., New York, McGraw-Hill Book Co., 1961, pp. ix, 586, $7.50.

THORNDIKE, ROBERT L., and HAGEN, ELIZABETH, *Measurement and Evaluation in Psychology and Education,* 2nd ed., New York, John Wiley and Sons, 1961, pp. viii, 602, $7.25.

THORPE, LOUIS P., KATZ, BARNEY, and LEWIS, ROBERT T., *The Psychology of Abnormal Behavior: A Dynamic Approach,* 2nd ed., New York, Ronald Press Co., 1961, pp. xii, 677, $7.50.

TYLER, LEONA, *The Work of the Counselor,* 2nd ed., New York, Appleton-Century-Crofts, Inc., 1961, pp. xiv, 327, $4.75.

VERNON, PHILLIP E., *The Measurement of Abilities* (reprint of 2nd ed.), New York, Philosophical Library, Inc., 1961, pp. vii, 276, $7.50.

WICKENS, DELOS D., and MEYER, DONALD R., *Psychology,* rev. ed., New York, Holt, Rinehart and Winston, 1961, pp. ix, 771, $7.75.

WIENER, NORBERT, *Cybernetics or Control and Communication in the Animal and the Machine,* 2nd ed., New York, John Wiley and Sons, 1961, pp. xvi, 212, $6.50.

ZULLINGER, HANS, Der Zullinger-Taflen-Test, 2nd ed. Stuttgart, Verlag Hans Huber, 1962, pp. 290, Fr/DM 28.50.

SOFT-COVERED REPRINTS

(Dates of original publication given in parenthesis is available).

CAPRIO, FRANK S., *The Power of Sex,* New York, Citadel Press, 1962 (1952), pp. 228, $1.50.

COUÉ, ÉMILE, and BROOKS, C. H., *Better and Better Every Day,* New York, Barnes and Noble, Inc. (Unwin Books), 1961, pp. 158, 95¢.

FREUD, S., *A Young Girl's Diary,* New York, Barnes and Noble, Inc. (Unwin Books), 1961 (1919), pp. 175, 95¢.

GILBERT, G. M., *Nuremberg Diary,* New York, New American Library, 1961 (1947), pp. 430, 75¢.

GRINKER, ROY R. *Psychosomatic Research,* New York, Grove Press, Inc., 1961, (1953), pp. 221, $1.95.

GRODDECK, GEORG, *The Book of the It,* New American Library, 1961, pp. x, 256, 75¢.

GUTHEIL, EMIL A., *The Handbook of Dream Analysis,* New York, Grove Press, Inc., 1960 (1951), pp. xv, 710, $2.95.

KLEIN, MELANIE, *The Psychoanalysis of Children,* New York, Grove Press, Inc., 1960 (1932), pp. 399, $2.45.

LILLIE, WILLIAM, *An Introduction to Ethics,* New York, Barnes and Noble, Inc., 1961 (1948), pp. ix, 342, $1.25.

LURIA, A. A., *The Nature of Human Conflicts: An Objective Study of Disorganization and Control of Human Behavior,* New York, Grove Press, Inc., 1960 (1932), pp. xvi, 431, $2.45.

PAVLOV, I. P., *Conditioned Reflexes,* New York, Dover Publications, Inc., 1961 (1926), pp. xv, 430, $2.25.

SALTER, ANDREW, *Conditioned Reflex Therapy,* New York, G. P. Putnam's Sons, 1961 (1949), pp. x, 359, $1.75.

STEBBING, L. SUSAN, *A Modern Elementary Logic,* New York, Barnes and Noble, Inc., 1961 (1943), pp. x, 214, $1.25.

STEKEL, WILHELM, *Auto-erotism: A Psychiatric Study of Onanism and Neurosis,* New York, Grove Press, Inc., 1961 (1950), pp. vii, 289, $2.95.

TASHMAN, HARRY F., *Today's Neurotic Family,* New York, University Publishers, Inc., 1960 (1957), pp. vii, 214, $1.65.

TAYLOR, A. E., *Elements of Metaphysics,* New York, Barnes and Noble, Inc., 1961 (1903), pp. xxvi, 419, $1.95.

THOULESS, ROBERT, *The Psychology of Religion,* Cambridge, Eng., University Press, 1961 (1923), pp. xxiv, 286, $1.75.

ZAEHNER, R. C., *Mysticism, Sacred and Profane,* New York, Oxford University Press, 1961 (1957), pp. xiv, 256, $1.50.

INDEX

CHARLES VALLEY BROOK, University of Texas

AUTHORS

(The names of authors of original articles are printed in CAPITALS and SMALL CAPITALS; of authors of books reviewed, in roman; and of reviewers, in *italics*.)

Ahmann, J. S. 343
Alexander, H. 710
ALLUISI, E. A. 674
Andreas, B. G. 170
ANDREWS, J. 667
Angona, L. 706
Applezweig, M. H. . 701
ARNOULT, M. D. . 675
Atkinson, R. C. ... 707
AXELROD, S. 673

Bailey, J. C. 520
BAKAN, P. 302, 457
Baker, L. M. 170
Baldwin, A. L. 342
Barnes, L. B. 520
BARRATT, E. S. 324
Barron, M. L. 516
BARTZ, A. E. 657
BAXTER, J. C. 66
BECKER, S. W. 115
BEHREND, E. R. ... 18
Berg, I. A. 708
Berlyne, D. E. 332
Berry, C. A. 167
BEVAN, W. 612
Birney, R. C. 523
BITTERMAN, M. E. 18, 421, 634
BOLLES, R. C. 699
Bonner, H. 701
BORING, E. G. ..326, 327
BRAUNSTEIN, M. L. 309
Brengelmann, J. C. . 176
BRINDLEY, G. S. ... 172
BROGDEN, W. J. ... 155
BROOKS, V. 624
BROWN, D. R. 144
Brown, F. 174
Brown, J. S. 699
Buchan, A. H. 169
BUCKHOUT, R. 429
BUGELSKI, B. R. ... 665
Bühler, K. 338
BULLOCK, D. H. .. 634

Caldwell, W. E. ... 705
CAMPBELL, R. A. .. 401
CARDNO, J. A. 499
CARLSON, V. R. 462, 644
CARRAN, A. B. 466
CAZAVELAN, J. 140
Chapanis, A. 706
CHAPLIN, M. R. ... 495
Clark, J. V. 520
CLARK, M. 324
Clarke, A. D. B. ... 340
Clarke, A. M. 340
Clarke, M. G. 174
Colby, K. M. 345
Coleman, J. C. 708
COLLINS, C. C. 470
Cooper, R. 334
CORNSWEET, T. N. 485, 653
CROVITZ, H. F. 271
CROWDER, W. F. .. 306

DALLENBACH, K. M. 676
Dalton, M. 165
DATTA, L.-E. G. ... 531
David, H. P. 176
Dawson, J. G. 710
Deatherage, B. H. .. 707
Dellis, N. P. 710
DeSilva, E. R. 698
DEWOLFE, A. S. .. 265
DIGGORY, J. C. 94
Diggory, J. C. 696
Doob, L. 350
Dorcus, R. M. 698
DUFORT, R. H. 661
DUNCAN, C. P. ..77, 242
DYER, D. W. 56

EBENHOLTZ, S. 193
EDWARDS, W. 311
EPSTEIN, W. 121
Estabrooks. G. H. . .344
EYSENCK, H. J. 127
Eysenck, H. J. 510

FARRIMOND, T. 477
FARROW, B. J. 671
Feld, S. 351
Field, J. 693
Field, M. J. 341
Finn, M. H. P. 174
FISHER, G. H. 321
FLOCK, H. 501
Fouraker, L. E. 347
Frank, J. D. 512
Freeman, E. 338
Freeman, F. S. 349
Freiberg, A. C. 522
Fruchter, B. 707

GALANTER, E. 208
Galanter, E. 161
GARVIN, E. A. 259
GIBSON, E. J. 554
GIBSON, J. J. 501
GOLDSTEIN, A. G. .. 667
GONZALEZ, R. C. .. 421
Gonzalez, R. C. 330
GREEN, E. J. 294
GRISWOLD, B. J. ... 35
Gubin, G. 351

Hallman, R. J. 169
HAMMOND, M. 554
HARCUM, E. R. 56
Harding, J. 351
HARTMAN, A. 587
Havighurst, R. J. .. 334
HELM, C. E. 437
HOCHBERG, J. 624
HOGAN, J. A. 307
Hook, S. 518
HUFF, E. M. 665
HUNT, W. A. 506
Hurvich, L. 172

Iacono, G. 706
Inhelder, B. 342
IVERSON, M. A. ... 504

Jackson, D. D. 514
Jacobson, E. 706

JAHNKE, P. K. 445
Jameson, D. 172
JENSEN, A. R. 390, 470
JOHNSON, R. 284
JOHNSON, R. E. 629
JONES, F. N. 147, 669

KAPPAUF, W. E. 482
Katona, G. 522
Katz, R. L. 520
KENDLER, H. H. 571
KENDLER, T. S. 571
Kennedy, J. L. 328
Kennedy, J. L. 347, 709
KEPPEL, G. 1, 353
KINKADE, R. G. 221
Kutner, B. 702

L'ABATE, L. 251
Lambert, E. C. 518
Lambert, W. W. 704
Lawrence, P. R. 520
Layton, W. L. 350
LEARNARD, B. 571
Ledwith, N. H. 349
Leeper, R. W. 336
LEPLEY, W. M. 310
LEVIALDI, A. 325
LEWIS, D. J. 77
Liddell, H. S. 693
Lifton, W. M. 351
LIGHTFOOT, C. 503
Lilienthal, J. 334
LIPTON, J. 324
Littman, R. 346
Lordahl, D. L. 352
Lucas, A. 706
LUCE, R. D. 35

Mackworth, N. 706
MacLeod, R. C. 709
Madison, P. 345
Mann, F. C. 165
Marcuse, F. L. 352
Matarazzo, J. D. 512
MCDONALD, R. P. 152, 619
McFarland, R. A. 167
McGlothlin, W. H. 698
McGuigan, F. J. 352
MCNEILL, D. 316, 492
Meares, A. 344
MEDLIN, R. E. 319
Messick, S. 176
MEYER, M. F. 323
MICHELS, K. C. 144
Miller, G. A. 161
Miller, G. E. 702
Miller, J. G. 348
MILLER, W. C. 505
Millman, J. 710
MILSTEIN, S. 421
Misiak, H. 709
Moodie, W. 352
MOORE, M. E. 289
More, D. 334
Mowrer, O. H. 330, 332
MUNSON, R. F. 640
Murphy, G. 163, 709
MYERS, L. B. 457

NAYLOR, J. C. 144
NEWMAN, S. E. 102, 429

Orth, C. D., III 520
OSSER, H. 554
OSTROFF, B. 94
OYAMA, T. 45

PEACOCK, L. J. 648
Peck, R. F. 334
Pepper, S. C. 696
Peters, H. N. 510
Piaget, J. 342
PICK, A. 554
Pick, H. 173
POFFENBERGER, A. T. 677
POLLACK, R. H. 495
POSER, E. G. 304
POSTMAN, L. 372
POULTON, E. C. 109
Pribram, K. H. 161
PRITCHARD, J. F. 612
Pronko, N. H. 350

RAAB, D. H. 298
REED, W. G. 612
REHULA, R. 353
Remmers, H. H. 334
Rethlingshafer, D. A. 705
ROBERTS, W. A. 134
ROCK, I. 193
ROSENZWEIG, M. R. 316, 492
ROYCE, J. R. 221
ROZIN, P. 307
Ryan, T. A. 161, 345, 523

SALTZ, E. 102
SANTOS, J. F. 259, 671
SANUA, V. D. 328
SCHOONARD, J. 457
SCOTT, T. R. 319
Sebeok, T. A. 351
Segall, M. H. 341
Seiler, J. A. 520
Sells, S. B. 167
Sheer, D. E. 707
Shock, N. W. 516
SIEGEL, S. 115
Siegel, S. 347
Silver, R. J. 710
SINGER, D. 147
SLATER, P. 127
Sloan, W. 340
SMALL, A. M., JR. 401
Smith, F. V. 336
Smith, P. C. 706
Solley, C. M. 163
SQUIER, R. W., JR. 294
STAYTON, W. R. 221
STEVENS, S. S. 596
Stewart, L. H. 350
Stone, H. K. 710
Suci, G. J. 351
Sulzer, E. S. 351
Sundberg, N. D. 514
Suppes, P. 707
Swanson, G. E. 704
Szeminska, A. 342

TASSONE, E. P. 644
Teevan, R. C. 523
TERRACE, H. S. 596
THOMPSON, L. 673
THOMPSON, R. 302
THOMSON, C. 284
TOCH, H. H. 605
Toman, W. 346
TREHUB, A. 140
Triades, H. C. 520
TUCKER, I. 140
TUCKER, L. R. 437
TUNNELL, E. H. B. 690
Turner, A. N. 520
TWELKER, P. A. 147

Uhr, L. 348
UMEMOTO, T. 85
UNDERWOOD, B. J. 1, 353
UTTAL, W. R. 150

Vernon, P. E. 343
Veroff, J. 351
VREELAND, R. W. 470

Walk, R. D. 170
WALLACH, H. 289
WALLACH, L. 277
WARREN, R. M. 109, 452
Waters, R. H. 705
WAUGH, N. C. 177
Wendt, G. R. 348
WHITE, B. W. 411
Whyte, L. L. 345
WILLETT, R. A. 127
WILLIAMS, M. 648
Winn, R. B. 173
Wohlwill, J. F. 163
Wolman, B. B. 336
WOSKOW, M. H. 669
WRIGHT, B. 232

YOUNG, P. T. 482

ZENER, K. 271
Zimmermann, R. R. 705

SUBJECTS

(References in *italic* figures are to reviews.)

Abnormal psychology, handbook of, *510 ff.*
Acculturation, progress toward, *350.*
Acknowledgment, an, 676.
Acquisition, conditioned responses, verbal, 265-270.
Activity recorder, ultrasonic, 648-652.
Aerial perspective, 501-503.
Adaptation-level, loudness judgments and, 612-618.
Affective scale, happiness, 211 f.; unhappiness, 213 f.
After-effects: kinesthetic, inspection and, 457-461, pre-inspectional judgments and, 202 f.; movement from rotating spiral, determinants of duration of, 127-133; spiral, apparatus, 310, psychophysical measurement of, 319-321; visual, figural, decay of, 242-250.
After-images, size, apparent, normal vs. schizophrenic *Ss*, 587-595.
Algesimeter, 304 f.
Animals: Earthworm, 531-553.
  Fish, conditioning in, 421-428; improved lever for, 307 f.
  Goldfish, avoidance-conditioning, 18-34.
  Pigeon, probability-matching, 634-639.
  Rat, taste-preferences of, 482-484; threshold, auditory, 652.
Anxiety, learning and, 251-258; size-constancy, after-images and, 558, 592.
Apparatus: Activity-recorder, ultrasonic device, an, 648-652
  After-image, cue-reduction box, 589.
  Algesimeter 304 f.
  Autokinetic movement, 226-228.
  Automatic, for method of limits, 309.
  Binocular, depth-perception, 619 f.; rivalry, 310.
  Box for taste-preferences, rats, 482-484.
  Discriminative, 574.
  IBM punch-cards in research, 150-151.
  Learning, human, 470-476.
  Lip-reading test, a, 477-481.
  Loudness-judgments, 613 f.
  Maze, earthworm, 533 ff.
  Mechanical lever (fish), 307 f.
  Retinal image stabilized, 653-656.
  Shocking circuit, constant current, 306 f.
  Statistics, sampling device, 657-660.
  Stereoscopic, new, 411-420.
  Thermal stimulator, waterless, 147-149.
  Wake Forest College laboratory, 661-664.
Apprehension, range of, pronounceability and, 560-568.
Archives of Psychology, status of, 686 f.
Aspiration, level of, utility and, 115-120.
Associations, verbal, size-distant, 644-647.
Associative value, factor in learning, 251-258.
Attainment tests, intelligence and, *343 f.*
Attention, position, determinant of, 56-65.
Attitude, consumer, *522 f.;* perspective, size-constancy and, 644-647.
Attributes, auditory, volume, 596-604.
Audition, loudness-judgments, 452-456; pitch, illusion, 323 f.; threshold, temporal differential, 401-410; volume, quantification of, 596-604.
Auditory, reaction-time, stimulus-duration and, 294-297; threshold, rat, 652.
Autokinetic, illusion, meaning, effect of, on, 605-611; movement, reduction of, 211-231.
Autism, perceptual effect of, 263.
Automation, *706.*
Autophonic judgments, loudness of, 452-456.

Behavior, drugs and, 348 f.; effective, *708 f.;* learning theory and, *330 ff.;* motivation in, *697 ff.;* neuro-, integrative systems, *707 f.;* pathological, Ghana, *341 f.;* plans and structure of, *520* *161 ff.;* purposive, *693 f.*, social, *520* *ff.;* systems of, *336 ff.*
Beliefs, primitive, *704 f.*
Betting, horse race, *698 f.*
Bibliography, Woodworth, R. S. 690-692.
Binocular rivalry, 310, vision, monocular vs., 56-65.
Biography: Bentley, M., 156 f.; Carr, H., 155; Cattell, J. McK., *679 ff.;* 682; Garmann, C. E., 678; Ladd, G. T., 683; Lashley, K. S., 506 ff.; McDougall, W., 684; Münsterberg, H., 685; Schäfer, E. H., *679;* Sherrington, C., 680; Thorndike, E. L., *679* f.; Titchener, E. B., 684 f.; Watson, J. B., 684 f. *See* Necrology.
Books received, *525-528; 711-714.*

Brain, electrical stimulation of, *707 f.*

Brightness, discrimination, children, size vs., 571-586; effect on size-illusion, 45-55.

Brunswik ratio, artifact of, 152-154.

Business psychology, bargaining, *347 f.*

Causality, phenomenal, 321 ff.

Character development *334 ff.*

Children, discrimination, brightness vs. size, 571-586; geometrical concepts, *342 f.;* longitudinal study (Rorschach), *349;* recognition, pictorial, 624-628.

Choice, gambling situations, in, 640-643, 698 f.; games of chance, 35-44; utility and subjective probability, 208-220.

Civilization, progress toward, *350.*

Clinical psychology, training for, *174 f.*

Closure, tested, 629-633.

Cognition, microgenesis and, 495-499; units of, 277-283.

Color, contextual influence on, 503-504; apparent warmth, weight and, 232-241; yellow, size-illusion and, 51 f.

Color-scales, structure of, 437-444.

Color-vision, problem, *523 f.*

Comparative psychology, *705.*

Complexity, concepts, of, 277-283.

Concept, attainment of, 277-283; geometrical, children's, *342 f.;* unconscious, of the, early history, *345.*

Concept-formation, discrimination, children, 571-586.

Conditioned reflex, history, *695.*

Conditioned responses, verbal, extinction, 265 f.

Conditioning; avoidance, extinction, 27-30; (goldfish), 18-34; classical (fish), 421-428.

Constancy, perceptual, index of, 152-154; shape-, 289-293.

Consumer, attitudes and expectations, *522 f.*

Contrast, lightness, 110 f.

Counseling, group and individual, *351 f.*

Critical flicker frequency, auditory discrimination and, 324 f.; intelligence vs., 324 f.

Culler, E. A. K., 155-160.

Cutaneous, galvanic skin response, 140-143; stimulator, waterless thermal, 147-149.

Decision-making, gambling, in, 640-643; group, *347 f.*

Depth-perception, binocular, 619-623.

Development, character, of, *334 ff.;* children (Rorschach study), *348.*

Discrimination, brightness vs. size, 571-586; 'dull' vs. 'bright' *S*s, 294-297.

Distance, cue of, 501-503.

Drugs, behavior and, *348 f.;* size-constancy judgments, effect of, on, 462-465.

Duration, differential threshold, conditioning, 401-410.

Earthworm, learning in, 531-553.

Emmert's law, schizophrenia, in, 587 f., 591 ff.

Error, time-order, loudness-judgments in, 612-618.

Expectation, gambling and, 35 f., 77 f.

Experimental psychology, *352;* texts in, *170 ff.*

Experimental upsurge, in psychology, 499-501.

Extinction, avoidance-conditioning, 27-30; conditioned verbal responses, 265-270; expectation and resistance to, 77-84.

Eye-dominance, effect on report, 56-65; test for, 271-276.

Fields of psychology: 1797-1874, 449 f.

Figural after-effect, *see* After-effects.

Fish, *see* Animals.

Fish-lever, improved, 307 f.

Fixation, visual, report and, 56-65.

Flicker, frequency, intelligence and, 324 f.

Forgetting, measuring, methods of, 1-17.

Form-perception, 495-499.

Galvanic skin response, components of, 140-143; semantic differential and the, 66-76.

Gambling, choices in, 35-44; betting, horse racing, *698 f.;* decision-making in, 640-643; expectation in, 77-84; probability, 35-44.

Geometrical concepts, children's, *342 f.*

Gerontology, social, America, *516 ff.*

Gestalt, closure tested, 629-633; psychology, principles of, *338 f.*

Goal, utility and, 115-120.

Gods, birth of, *704 f.*

Goldfish, *see* Animals.

Group counseling, individual and, *351 f.*

Hand-dominance, test for, 271-276.

Happiness-unhappiness, affective scale, 211-213.

Healing, persuasion and, *512 ff.*

History, conditioned reflex, *695;* physiological psychology in 1920s, *694 f.;* psychology, articles 1797-1874, 499-501; unconscious, early, of the, *345.*

Hue, effect on size-illusion, 45-55.

Hypnosis, medical, *344;* uses of, *352.*

- Illusion, apparent warmth and weight of colors, 232-241; autokinetic, 605-611; contextual influence on color, 503-504; pitch, new, 323 f.; size- (concentric circles), 45-55; size, effect of brightness and hue on, 45-55; spiral, rotating, 127-133, 310, 320; stroboscopic movement, 193-207.
- Image, behaviorism and the, *333;* retinal, stabilized, 653-656.
- Individual differences, apparatus, for study of, 470-476; color-perception, structure of, 437-444; lip-reading, 481; reproduction of binary stimuli, in, 56-65; success, in estimates of, 94-101.
- Industry, human relations in, *520 ff.*
- Information, measures of, 311-316.
- Instinct, problem, *523 f.*
- Intelligence, critical flicker frequency vs., 324 f.; tests vs. attainment, *343 f.*
- Intensity, loudness-judgments, 452-456.
- Interactions, multiperson, *707.*
- Interposition, depth-perception and, 619-623.
- Isolation, effect of, in serial learning, 134-139.

- Judgment, depth, binocular, 619-623; kinesthetic after-effects, pre- vs. post-inspectional, 202 f.; lightness, 109-114; loudness, basis of, 452-456, time-order, effect of, on, 612-618; pre-inspectional, kinesthetic after-effects and, 202 f.; ratio and partition, 109-114; size-constancy, drugs, effect of, on, 462-465.

- Kinesthetic after-effects, *see* After-effects.

- Laboratory, Wake Forest College, 661-664.
- Landis, Carney: 1897-1962, 507-509.
- Language, style in, *351.*
- Learning: Apparatus, human, 470-476.
  - Associative connection and, 6 f.
  - Avoidance (goldfish), 18-34.
  - Cultural and linguistic differences in, 85-93.
  - Curves of, 5 ff.
  - Discriminative, 294-297; children, brightness vs. size, 571-586.
  - Earthworm, 531-553.
  - Efficiency, set, effects of, on, 445-452.
  - Habit-reversal, 546 ff.
  - Human, apparatus for study, 470-476.
  - Incidental and intentional, 284-288.
  - Lip-reading, test of, 477-481.
  - Mathematical equation of, 177-192.
  - Neural basis of, *694.*
  - Paired-associate, 85-93; 102-108, item-selection in, 353-371, repetition and, 353-371; 466-469.
  - Perceptual, theory of, *163 ff.*
  - Rate, dull vs. bright *S*s 294-297.
  - S–R vs. R–S, 429-436.
  - Serial, isolation, effect of, in, 134-139, length of series and, 177-192, meaning, effect of, in, 134-139, temporal vs. spatial effects, 390-400.
  - Serial position, a cue, in, 102-108.
  - Set and, 445-451.
  - Similarity, a function of, 85-93.
  - Single-trial, 191 f.; 353-371; 372-389; 466-469.
  - Spatial effects in, 390-400.
  - Syntactic structure and, 121-126
  - Temporal effects in, 390-400.
  - Theory, behavior and, *330 ff.*, generalization, 429-436, symbolic process and, *332 ff.*
  - Transfer of, 251-258.
  - Variability, dull vs. bright *S*s, 294-297.
- Lightness, judgments of, 109-114.
- Limens, differential, auditory, 401-410.
- Lip-reading, test of, 477-481.
- Lorge and Thorndike, semantic count, 492 f.; inaccuracies, 316-319.
- Loudness, pitch and, estimates of, 669-671; time, effect of, on, 612-618.

- Magnitude, scale of, 669-671.
- Management, theory and feeling, *165 ff.*
- Massed practice, effects of, on spiral movement, 123-133.
- Maze, earthworm, 533 ff.
- Meaning, auditory, changes in, 673 f.; autokinetic illusion and, 605-611; test of Osgood's theory of, 66-76.
- Medical schools, teaching in, *702 f.*
- Medicine, aerospace, *167 ff.*
- Meetings: Amer. Philos. Soc., 327 f.
  - East. Psychol. Assn., 504 f.
  - Interamer. Soc. Psychol., 328 f.
  - Nat. Acad. Sciences, 326 f.
  - Rocky Mountain Psychol. Assn. 505 f.
  - Soc. Exper. Psychologists, 328.
  - Southern Soc. Philos. Psychol., 674 f.
  - Southwestern Psychol. Assn., 675 f.
- Memory-traces, qualitative changes in, 629-633.
- Mental deficiency, *340 f.*
- Mental health, national survey, *351.*
- Mental set, test revised, 665-667.
- Method, experimental and statistical, *352;* limits, automatic apparatus for, 309; machine collection of data, 150-

151; magnitude estimation, 596 ff.; paired-associate learning, 353-371; 372-389; 466-469; procedure, and, determining taste-preferences (rats), 482-484; psychophysics, single stimuli, 612-618, staircase, 485-486, scaling techniques, *see* Scaling; scientific, philosophy, psychoanalysis and, *518 ff.;* retention, evaluation of, 1-17.
Microgenesis, problems in, 495-499.
Mind-body problem, Woodworth's views, 687.
Models, learning, Markov, *707.*
Monocular vision, binocular vs., 56-65.
Motivation, *697 ff.;* effects of, on spiral movement, 127-133; learning and, 284-288; psychoanalytic theory of, *346 f.*
Movement, autokinetic, 221-231, determinants of, 608-611; illusory, rotating spiral, 123-133; 310; 320, visual, rotating spiral, 123-133; 310; 320; perception of, angle of approach and, 144-146; *phi,* 193-207; stroboscopic, basis of, 193-207.
Muscular effort, autophonic judgments and, 452-456.

Necrology: Culler, E. A. K., 155-160.
  Landis, C., 507-509.
  Woodworth, R. S., 677-689; bibliography, 690 ff.; portrait, 530.
Negative after-effect, rotating spiral, 127-133; 310-320.

Obituary, *see* Necrology.
Organizational life, business, *165 ff.*

Pain, apparatus for measuring threshold and tolerance-levels, 304; constant current shocking circuit, 306 f.
Paired-associates, method of, *see* Method.
Parapsychology, primer, *709.*
Perception: Aerial perspective, 501-503.
  Apparent distance, 501-503.
  Autism in, 263.
  Causality of, 321 ff.
  Color, contextual influence on, 503 f.
  Color-scales, 437-444.
  Constancy, index of, 152-154.
  Depth-, binocular, 619-623, stroboscopic, 411 f.
  Form, microgenesis of, 495-499.
  Integrating process, an, 325 f.
  Microgenesis and, 495-499.
  Movement, angle of approach and, 144-146; illusory, spiral, 127-133; 310-320.
  Random shapes, 667-669.
  Recency, effect of, on, 671-673.
  Reversible perspective vs. auditory changes in meaning, 673 f.
  Reward and punishment, effect of, on, 259-264.
  Shape, role of slant in, 289-293.
  Size-constancy, drugs, effect of, on, 462-465.
  Stereoscopic, new, a, 411-420.
  Temporal, audition, 401-410.
  Words, written, 554-570.
Perceptual world, development of the, *163 ff.*
Personality, *701 f.;* dynamics of, *708 f.;* research in, *176.*
Perspective, aerial, 501-503; attitude, size and the, 644-647; reversible, auditory meaning, changes in, and, 673 f.
Persuasion, healing through, *512 ff.*
Philosophical roots of psychology, *709 f.*
Philosophy, scientific method, psychoanalysis, and, *518 ff.*
Physiological psychology, *707 f.,* in the 1920s, *694 f.*
Physiology, handbook, *693-696.*
Pigeon, probability-matching, 634-639.
Pitch, illusion, new, 323 f.; loudness and, estimates of, 669-671.
Position, effect of, on report, 56-65; serial, a cue in learning, 102-108; temporal vs. spatial, learning, in, 390-400.
Practice, massed, effects of, on spiral movement, 123-133.
Probability, subjective, measurement of, 214-216, utility and, 208-220.
Probability-matching, pigeon, 634-639.
Problem, mental, water-jar, 665-667.
Projective test, Rorschach, longitudinal study (children), *349.*
Pronunciation, invariance of, 569 ff; rules for, 555.
Psychical research, *709.*
Psychoanalysis, scientific method, philosophy, and, *518 ff.*
Psychoanalytic, research, introduction to, *345 f.;* theory of motivation, *346 f.*
Psychology, aerospace, *167 ff.;* clinical, training for, *174 f.;* literature, of, *169 f.;* Soviet, symposium, *173 f.;* texts, experimental, *170 ff.;* theories and systems, contemporary, *336 ff.*
Psychosomatics, *696.*
Psychotherapy, comparative study of, *512 ff.;* criticism of, *511.*
Punishment, perceptual effect of, 259-264.
Reaction-time, auditory, intensity and,

294-297; stimulus-duration and, 294-297.
Reading, units of, 554.
Recency, effect on perception, 671-673.
Recognition, incomplete circles, of, 629-633; pictorial, unlearned, 624-628; word-, meaningfulness and, 558 f.; thresholds of, 557 f.
Reinforcement, problem, *523 f.*
Repetition, role of, learning, in, 353-371; 372-389; 466-469.
Report, position vs. accuracy of, 56-65.
Responses, mediated, 3–10 yr. old children, of, 571-586.
Retention, earthworms, 545 f.; 11 mo., after, 14 f.; methods of studying, 1-17; rehearsal, effect of, on, 4 f., 10 f.
Retina, visual pathways and the, *172 f.*
Retinal images, stabilization of, 653-656.
Reward, perceptual effect of, 259-264.
Rivalry, binocular, apparatus for, 310.
Rorschach test, longitudinal study of, *349.*
Russian psychology, symposium, *173 f.*

Sampling, devices for, 657-660.
Scales, category vs. ratio, 113.
Scaling, problem of, 210-214.
Schafer-Murphy effect, reëxamined, 259-264.
Schizophrenia, delusional vs. non-delusional, size of after-images, 587-595; etiology of, *514 ff.*
Schizophrenics, after-images of, apparent size of, 587-595; psychotherapy with, *710.*
Scientific psychology, philosophical roots, *709 f.*
Semantic, count, existing, described, 492; differential performance, 66-76. Lorge and Thorndike's, 492, inaccuracies, 316-319; similarity between S and R, 85-93; uses of, 492-495.
Sensation, autokinetic, *see* Autokinetic.
Sensitivity, differential, temporal, auditory, 401-410.
Set, compound vs. positive and negative, 445-452; learning and, 445-451; mental, efficiency of, 665-667.
Sex-differences, learning, in, 251-253.
Shape, constancy, 289-293; random, perception of, 667-669.
Shocking circuit, constant current, 306 f.
Single-trial learning, 466-469; gradual acquisition, special care of, 469.
Size, apparent, after-images, normal vs. schizophrenic *Ss*, 587-595; discrimination, children, brightness vs., 571-586.
Size-constancy: After-images, anxiety and, 588; 592; normal vs. schizophrenic *Ss*, 587-595.
  Attitude, perspective and, 644-647.
  Judgments of, drugs, effect of, on, 462-465.
Size-distance, verbal associations, 644-647.
Skin, *see* Cutaneous.
Slant, role of, shape-perception, in, 289-293.
Social, behavior, *520 ff.;* psychology, Ghana, rural, *341 f.*
Sound, attributes of, 596-604.
Soviet psychology, symposium, *173 f.*
Spatial, temporal vs., effects, learning, in, 390-400.
Spiral movement, after-effects of, 127-133; apparatus, 310; measurement of, 320.
Statistics, elements of, *710.*
Stereoscopic effect, dot matrices, 411-420; shape and slant, 289-293.
Stimulation, intra- and inter-sensory, 321 ff.
Stimulus-duration, reaction-time and, 294-297.
Stroboscopic movement, basis of, 193-207.
Strong Vocational Interest Blank, uses of, *350.*
Style, language, in, *351.*
Subliminal signals, 325 f.
Success, estimated, determinants of, 94-101; performance, a function of, 94-101.
Superstition, origin of, *704 f.*
Syntactic structure, learning and, 121-126.
Systems, behavioral, *336 ff.;* Gestalt, *338 f.;* psychological, contemporary, *336 f.*

Taste, apparatus, for determining rats' preferences, 482-484.
Teaching, medical schools, in, *702 f.*
Technology, automation, *706;* human adjustment and, *706.*
Temporal, spatial vs., effects, learning, in, 390-400.
Test, eye- and hand-dominance, 271-276; lip-reading, 477-481.
Tests, intelligence and attainment, *343 f.*
Theories, systems and, contemporary, *336 ff.*
Thermal, stimulator, *see* Apparatus.
Threshold, auditory, rat, 652; differential, auditory, temporal, 401-410; movement, 144-146; pain, apparatus for measuring, 304 f.; recognitive, words, 557 ff.

Time, filled vs. unfilled, 401-410.
Tonal, volume, 596-609.
Tragedy, psychology of, *169 f.*
Training, transfer of, 251-258; 680.

Uncertainties, freedom, degrees of, for, 312.
Uncertainty, problem of, 312-316.
Unconscious, the, before Freud, *345;* concept, early history, *345.*

Value, sources of, 696 ff.
Verbal, associations, size-distant, 644-647; behavior, semantic count in studies of, 492-495; responses, conditioned, extinction of, 265-270.
Vision, after-images, normal vs schizophrenic *S*s, of, 587-595; binocular vs. monocular, 56-65; color, problem of, *523 f.;* color-scales, 437-444; figural after-effect, amount and decay of, 242-250; hemispherical dominance, 56-65; image, retinal, stabilized, 653-656; perception of movement, 144-146; physiology of, *172 f.;* rotating spiral, 123-153; 310; 320.
Vocation, Strong Interest Blank, uses of, *350.*
Volume, tonal, quantification of, 596-604.

Woodworth, Robert Sessions, 677-689; bibliography, 690 ff.; portrait, 530.
Word-association, semantic count and, 494 f.
Words, perception of, 554-570.

Yellow, size-illusion and, 51 f.

BUREAU OF EDNL. & PSYL. RESEARCH LIBRARY